Tutorial IEDM 2015:

Advance Device Concepts for 7nm Node and Beyond

12/05/2015

Objective

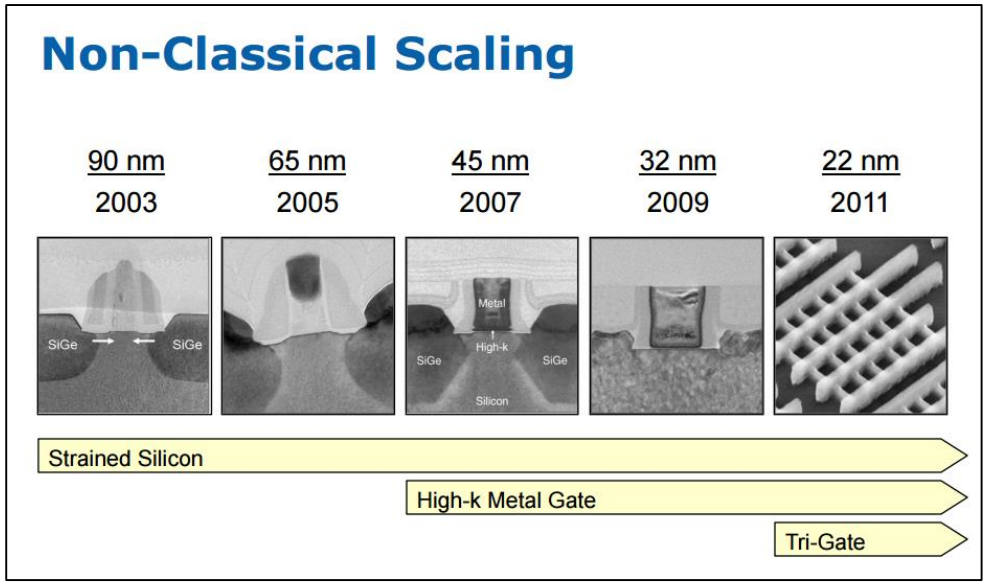
- Understand key device advantages and limitations of the non-ideal bulk FinFETs adopted for 14/16/22nm nodes
  - + I_{ON}, DIBL, SS, σV_T
  - - minimal L_G scaling
- Understand simple physical picture of nanoscale MOSFETs
  - Source-side injection MOSFET model
  - Mobility, velocity saturation, ballistic transport, quantum capacitance, quantum confinement
  - Key: Small conductivity m* channel direction, large confinement mass, large DOS mass are all key
- With simple picture of nanoscale MOSFET and advantage and limitations of non-ideal bulk FinFET, understand how future nodes 10,7, 5nm might develop
  - Bulk Si FinFET for 10 and 7nm
  - Tall fins
  - Subfin Leakage fix

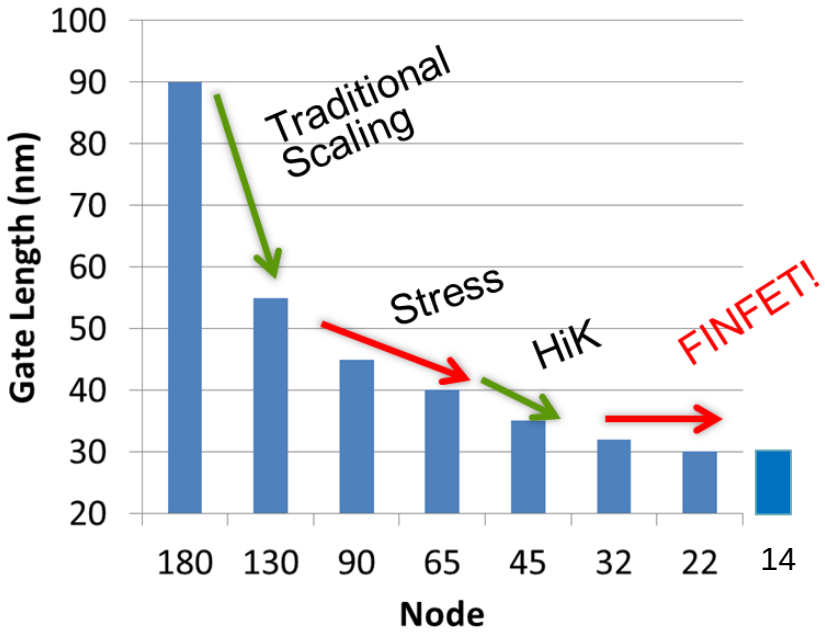
Outline

- **State-of-the-art for logic devices**
- FinFET device physics basics
- Metrics for advance logic devices
- Deeper look at some advanced device concepts
  - Essential Nanoscale Device Physics
    - mobility, ballistic transport, velocity saturation, thermal velocity, m* and density of states
  - Transistor variation / random doping effect
  - Strain I, Ge and III- channels
  - Quantum confinement
  - External resistance
  - Sub fin doping and gate all around devices
- Conclusion: How does the roadmap evolve?

Scaling Trends Past 2 Decades

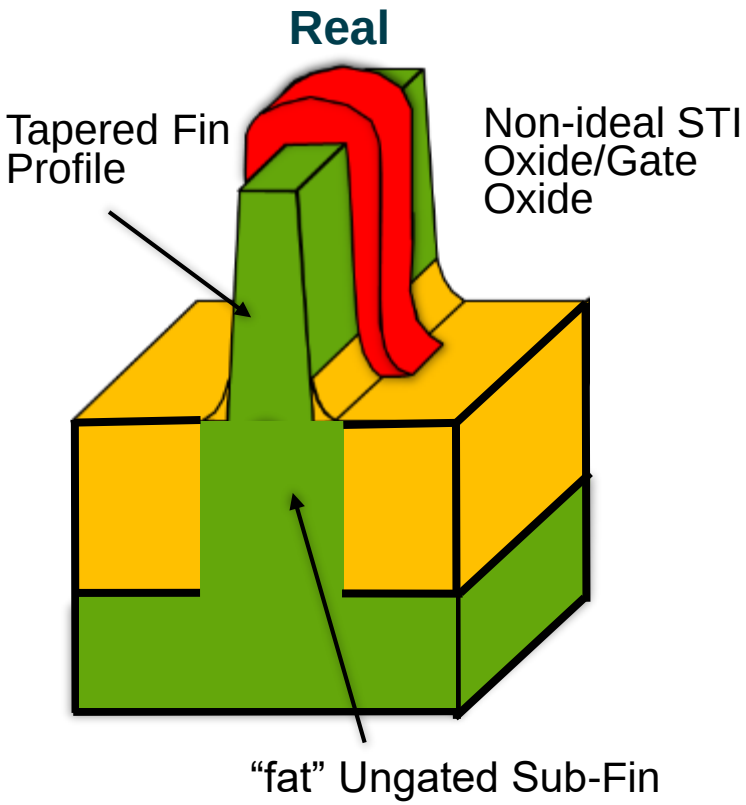
Source: Bohr / Intel

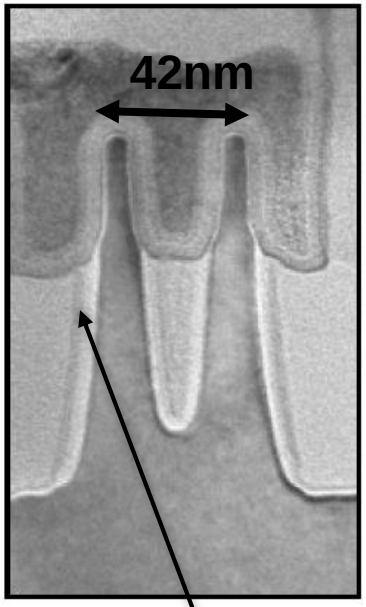




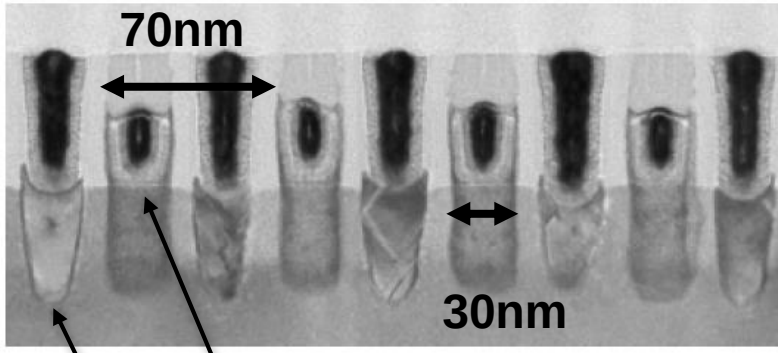
- Up until 130nm node, everything going according to plan (Dennard Scaling)
- T_{OX} scaling slows/stopped at 90-65nm
- T_{OX} replaced by mobility boost from stress – Moore's Law continues
- T_{OX} scaling continues with new materials but lasts maybe 2 generations
  - **L_G scaling has slowed this past decade even with introduction of high K and FinFETs and must be addressed for 10 and 7nm.**

State-of-the-Art: 14nm FinFET





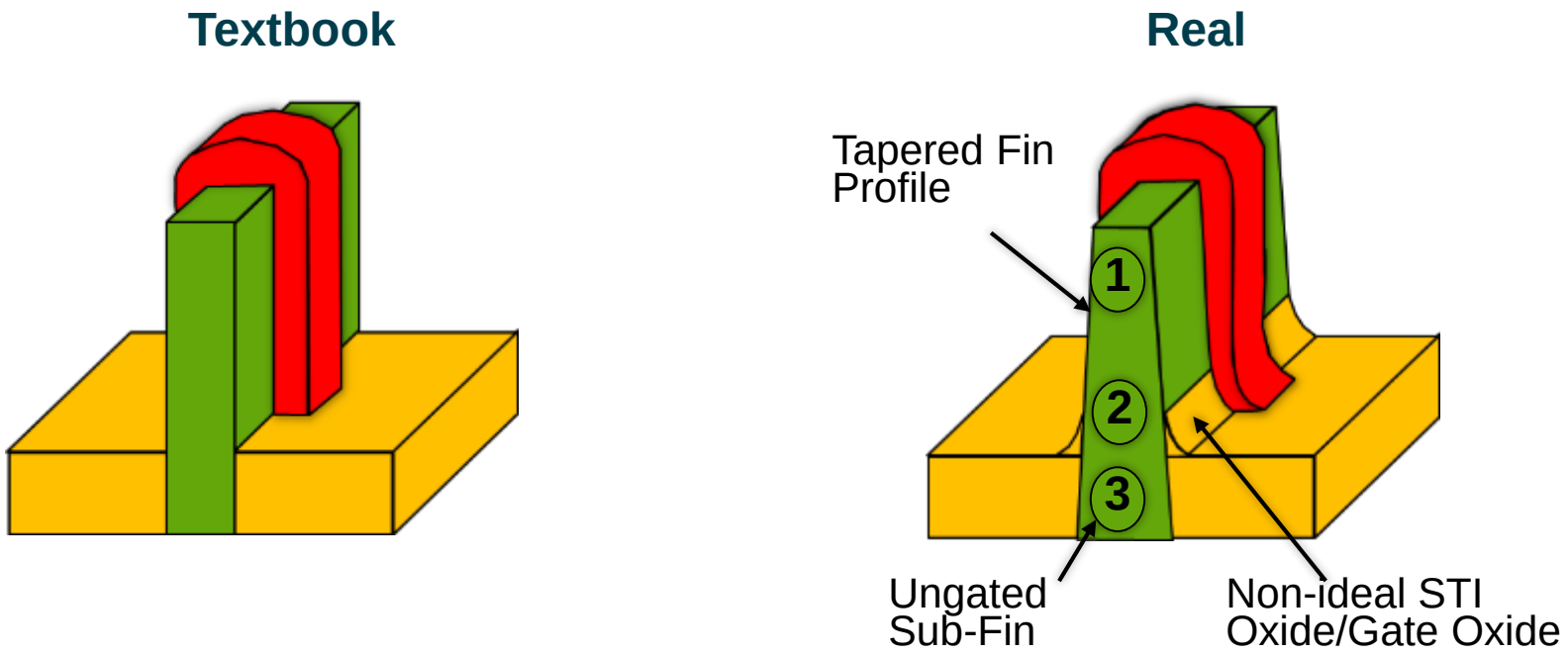
Intel 14 nm
Source: Chipworks



- 14 nm used for A9, Exynos 7420, Broadwell, and Snapdragon 820
- FinFET PLUS all the performance "boosters" of last 10 years
- Note: Bulk Finfet L_{GATE} ~30nm and will need to be scaled by 7nm

- **Bulk Finfet L_{GATE} ~30nm and will need to be scaled by 10 and 7nm**

Real vs. Textbook FinFET



- Non-uniformity in Fin etch and device architecture results in three distinct fin regions:
  1. Ideal “Tri-Gate” at the top, thin width, good gate control
  2. Wider FinFET in the middle, less gate control
  3. Sub-fin “BJT”, an ungated device which contributes to leakage

- **Sub-fin leakage needs to be address device as Si FinFET scaled to 10 and 7nm**

Outline

- State-of-the-art for logic devices
- **FinFET device physics basics**
- Metrics for advance logic devices
- Deeper look at some advanced device concepts
  - Essential Nanoscale Device Physics
    - mobility, ballistic transport, velocity saturation, thermal velocity, m* and density of states
  - Transistor variation / random doping effect
  - Strain I, Ge and III- channels
  - Quantum confinement
  - External resistance
  - Sub fin doping and gate all around devices
- Conclusion: How does the roadmap evolve?

Why FinFET: Scaling Requires Depleted Devices

$$DIBL = \frac{V_{Th}^{DD} - V_{Th}^{low}}{V_{DD} - V_{D}^{low}}$$

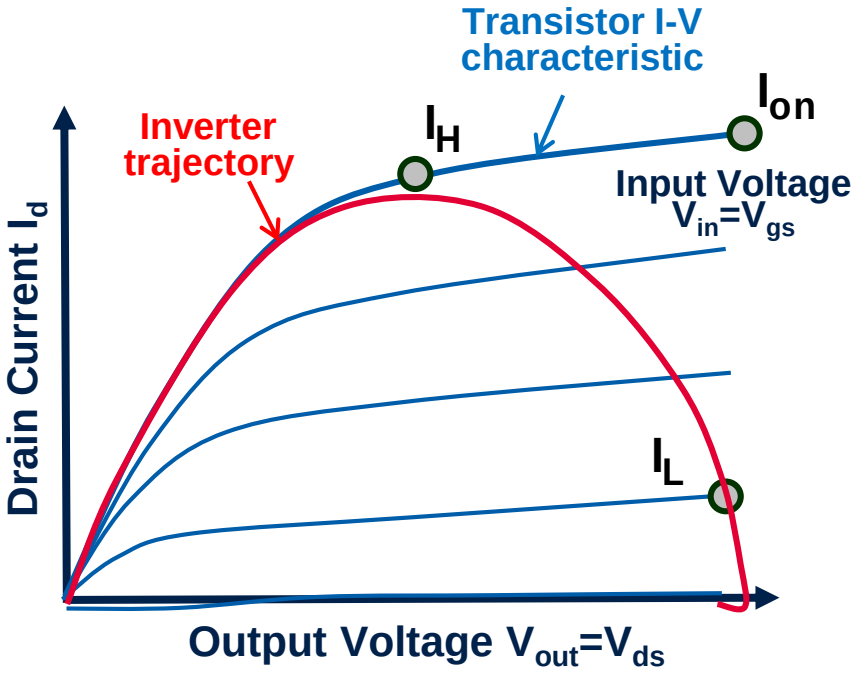
$$SS = \ln(10)\frac{kT}{q}\left(1 + \frac{C_{depl}}{C_{ox}}\right) \cong 60\frac{mV}{dec}\left(1 + \frac{C_{depl}}{C_{ox}}\right)$$

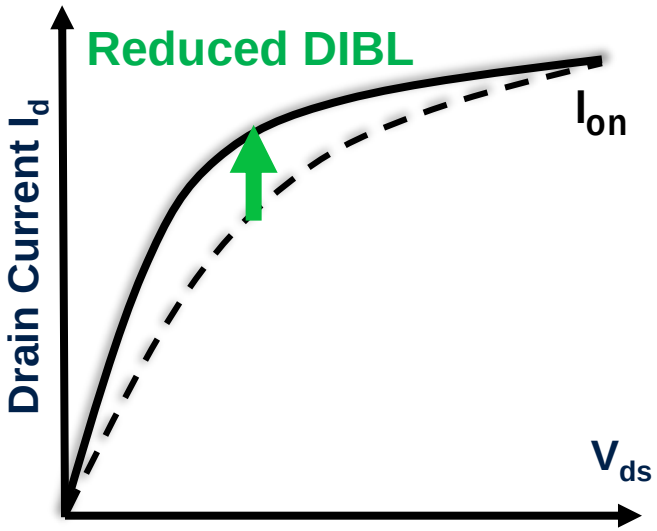
$$\sigma_{Vt} = \frac{qT_{ox}}{\varepsilon_{ox}}\sqrt{\frac{N_{sub}W_{dep}}{3LW}}$$

- Depleted devices improve SCE (both SS, DIBL, Ioff) and transistor matching
- Historically SCE addressed in Dennard classic scaling by increasing Cox (thinning Tox)
- Limited L_G scaling benefit due to drain to source sub-surface leakage path (non-gate controlled region) in FinFETs

- **Bulk FinFET is being designed for Electrical W, DIBL, SS, σVT benefit and this trend will continue for 10 and 7nm.**

Depletion also Improved Performance

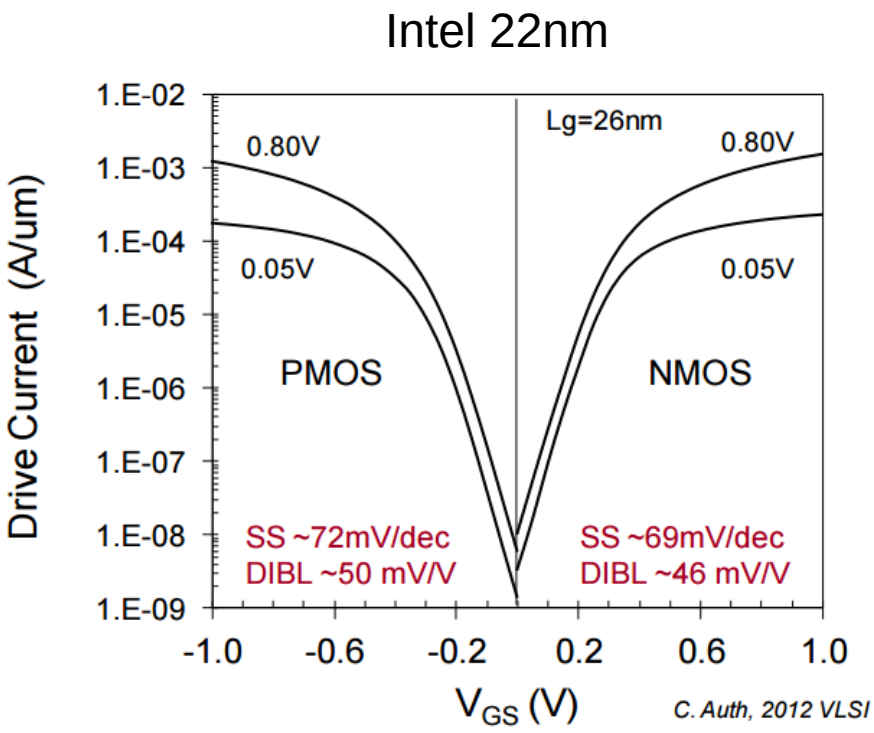


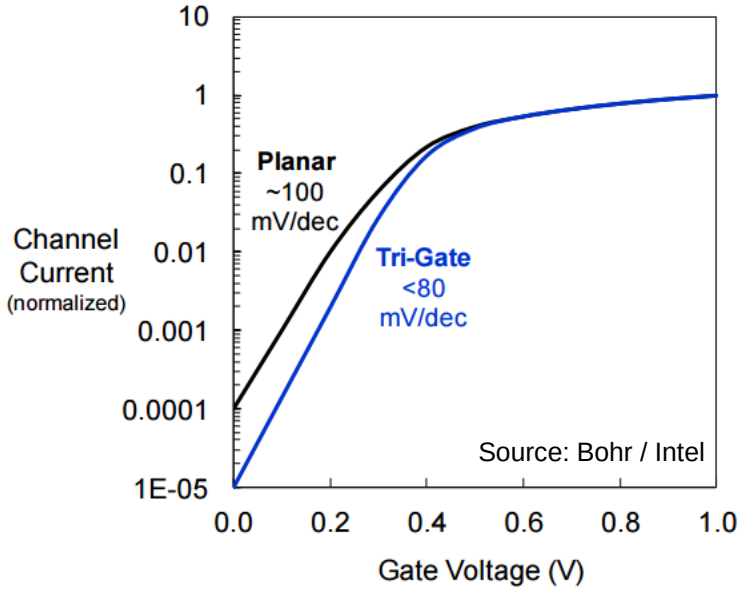


$$\text{Delay:} \quad t_d \propto \frac{C_{load} V_{dd}}{2 \cdot I_{eff}}$$

- Delay metric → Effective current : $I_{eff} = (I_H+I_L)/2$
- Undoped channels (FinFET, FD-SOI, DDC) have intrinsically low DIBL
  - This translates into higher I_{eff}, specially at low V_{dd}

Improved SS: Improved IOFF and/or Better Low V performance





$$SS = \ln(10)\frac{kT}{q}\left(1+\frac{C_{depl}}{C_{ox}}\right) \cong 60\frac{mV}{dec}\left(1+\frac{C_{depl}}{C_{ox}}\right)$$

- Fully depleted channel $C_{depl} \sim 0$ -→ SS ~ 60mV/dec

- **Improved SS Allows for Lower V_T for same I_{OFF} (key for advanced node VDD scaling)**

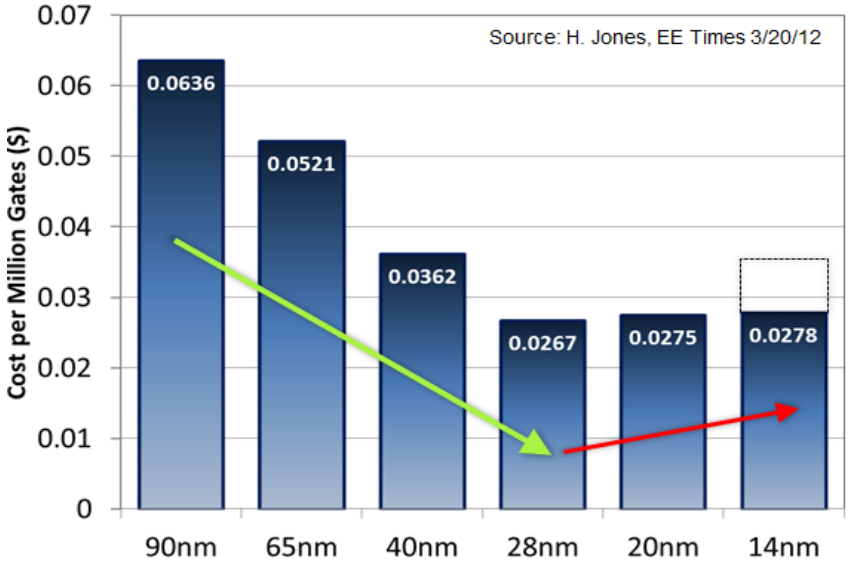
Outline

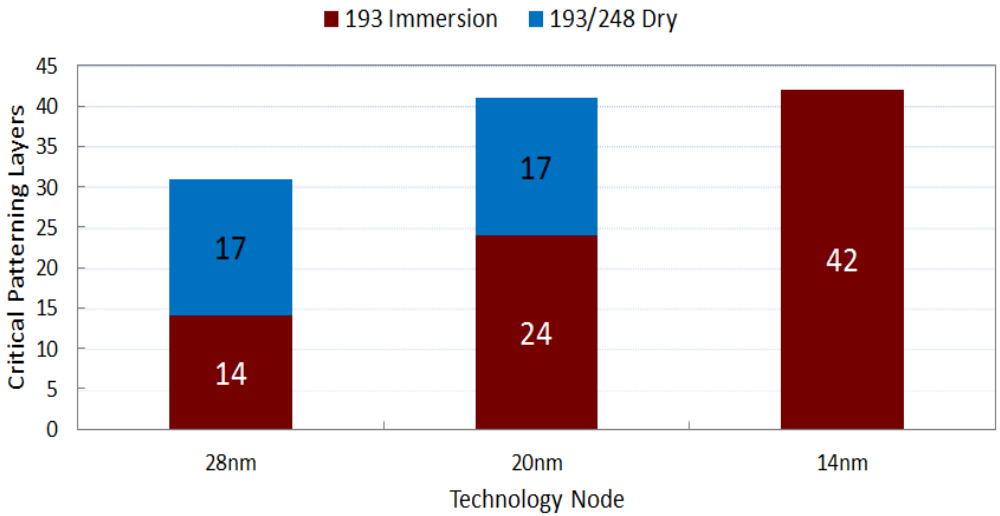
- State-of-the-art for logic devices
- FinFET Basics
- **Metrics for advance logic devices**
- Deeper look at some advanced device concepts
  - Transistor variation / random doping effect
  - Strain I, Ge and III- channels
  - Quantum confinement
  - Drive current and relation to mobility, velocity saturation, and density of states
  - External resistance
  - Sub fin doping and gate all around devices
- Conclusion: How does the roadmap evolve?

Executive Summary: Metrics Advanced Nodes: 14 to 7nm

Metric	Insight into Metric
Cost	• Moore's Law: Unlikely major adoption if not lower cost • 14,10,7nm patterning cost is slowing Moore's Law
I_D	• **Large mA/μm being targeted** • **Requires large electrical width in planar area** • **One of key FinFET strengths** • Helped by tall fins on tight pitch (increase fin parameter) • Note Intel's drive current is normalized by top down W not W_{EFF} • Large I_D being targeted even at some degradation to CV/I
DIBL	• Improved DIBL • Important for low voltage performance
SS	• Improved SS allows for lower V_T which enables lower V_{DD}
N/P ratio	• Close to 1 • **Allow pFET width to be similar to nFET (vs historical 2:1)** • Requires large uniaxial <110> stress on pFET • One of key reason FinFET on Bulk is adopted for 16 through 7nm
Transistor Matching	• Key for low voltage SRAM operation • Many sources of transistor variation • Fins are still doped at 14nm so RDF still a factor
L_{GATE}	• Smaller/ better BUT to date FinFET has only provided modest scaling

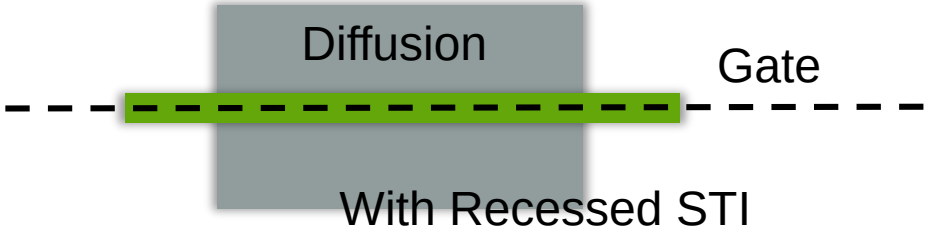
Transistor Cost Trend





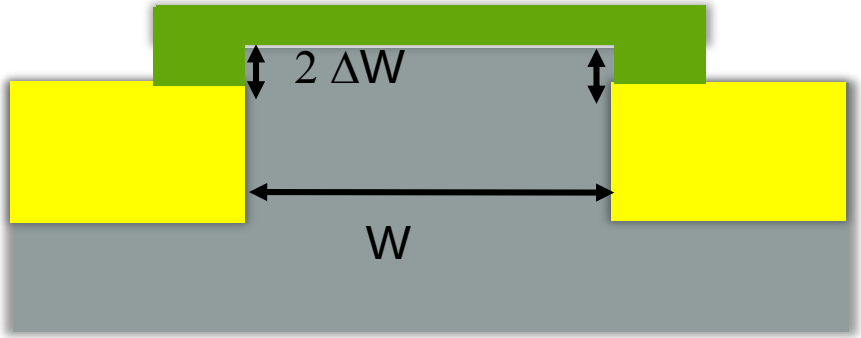
- Cost per transistor decreased until 28nm node then starts increasing
  - Mainly driven by lithography cost
- 20nm and below will be more expensive than 28nm for level of functionality

How to Think About FinFET: (Large $W_{ELECTRICAL}$)



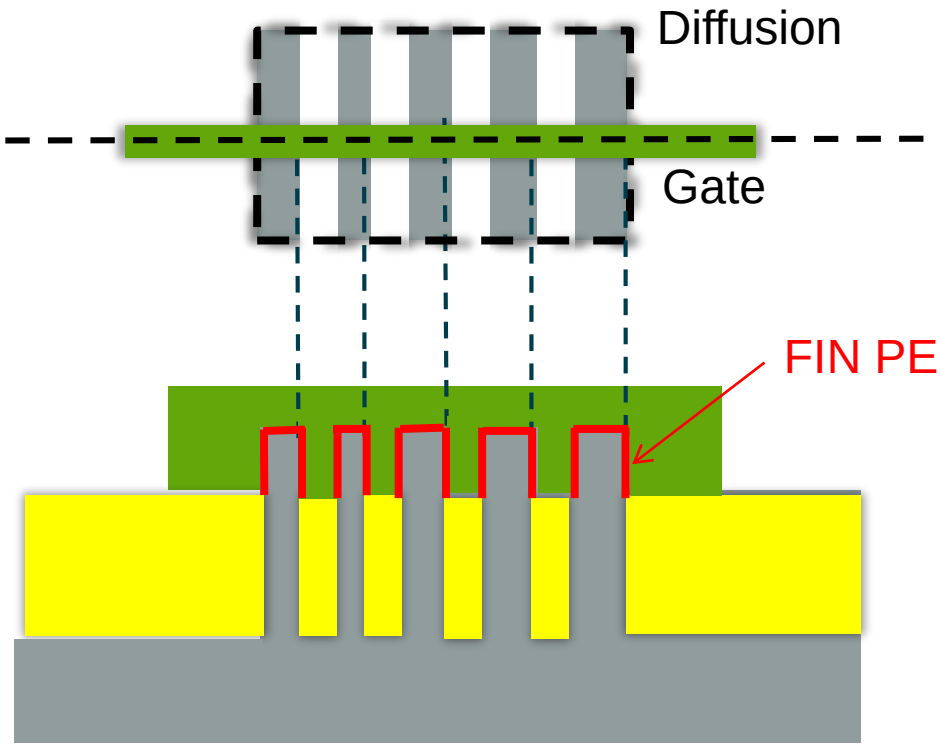
Top down planar transisor

Xsection planar transistor



$$W_{ELECTRICAL} = W + 2\,\Delta W$$

FinFET: Key Advantage is Large $W_{ELECTRICAL}$



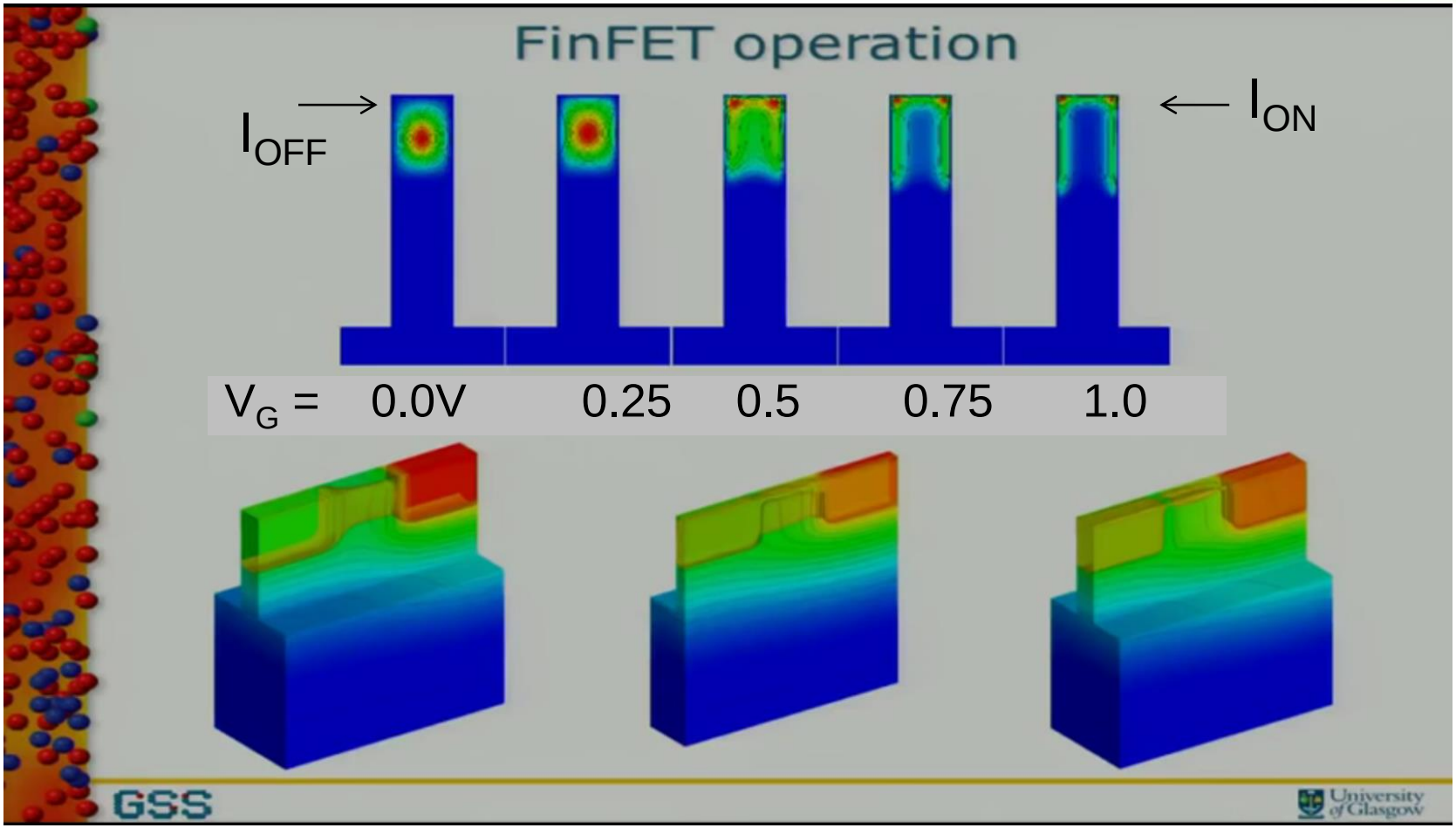
Cut Diffusion into strips

Recessed STI to set Fin Height

$$W_{ELECTRICAL} = FIN_{PERIMETER} = \#FINS\left(2FIN_H + FIN_W\right)$$

$$W_{ELECTRICAL} = \frac{PLANAR_{WIDTH}}{FIN_{PITCH}}\left(2FIN_H + FIN_W\right)$$

22/14nm Design Point: Little Volume Inversion



- **NO Significant bulk inversion in the fin bodies for 22/16/14 designs**
- **Will be issue as move to 10/7nm**

Normalize Currents by Planar W or Fin Parimeter?

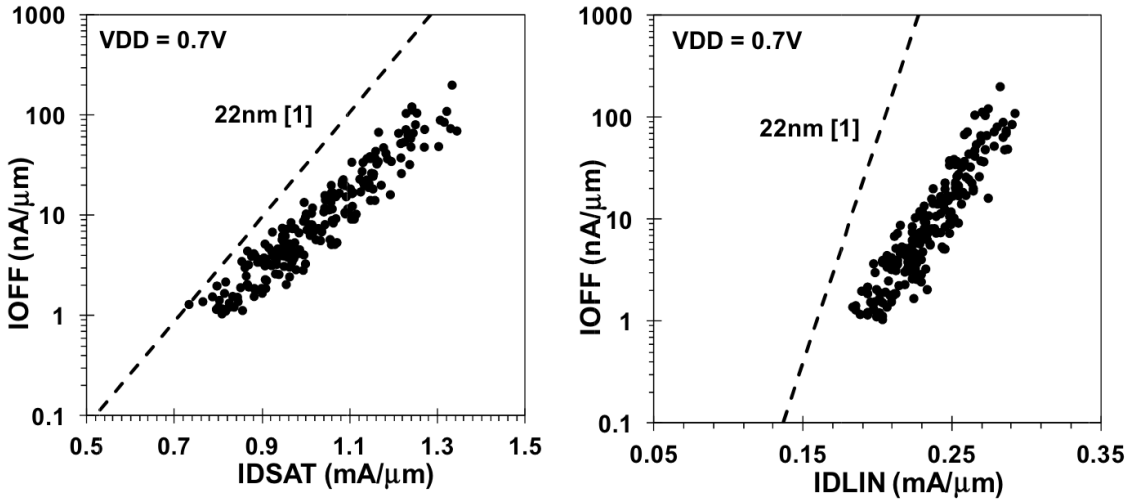
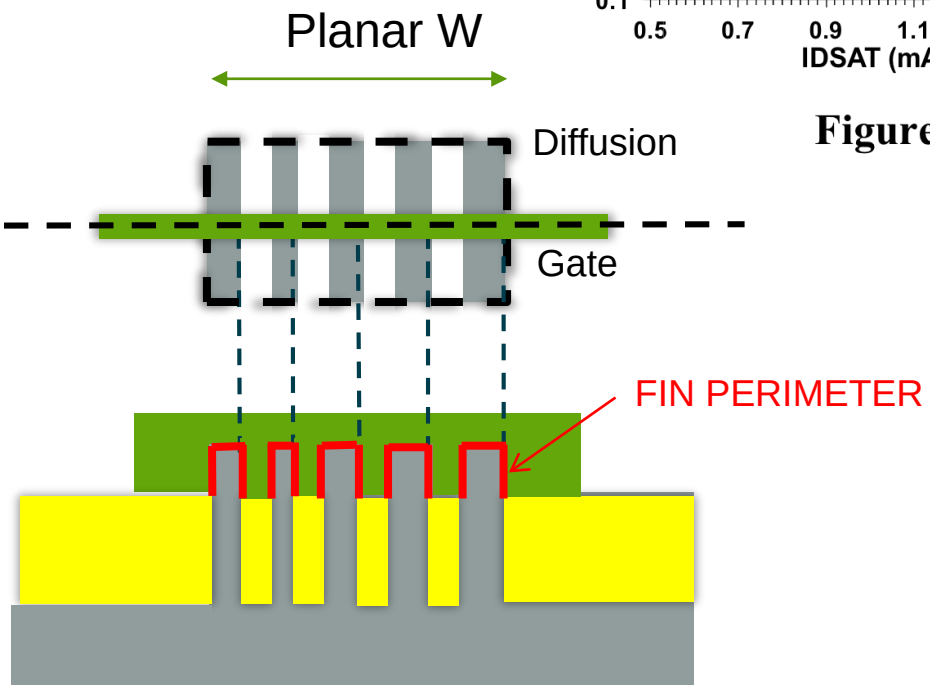


Figure 3: NMOS Idsat and Idlin curves



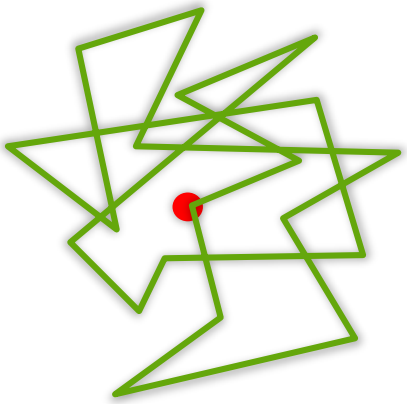
- **For Intel 14nm : (Fin perimeter / Planar W) = ($2Fin_H$ – Fin_W) / pitch ~2.2**

Outline

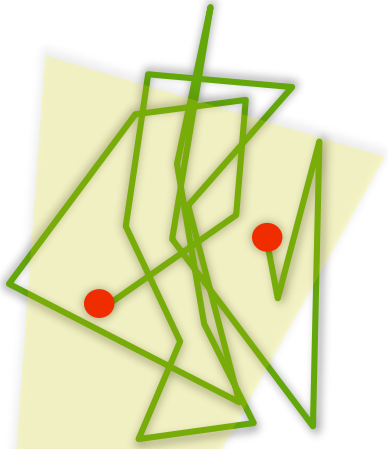
- State-of-the-art for logic devices
- FinFET Basics
- Metrics for advance logic devices
- **Deeper look at some advanced device concepts**
  - **Drive current and relation to mobility, velocity saturation, and density of states**
  - Transistor variation / random doping effect
  - Strain I, Ge and III- channels
  - Quantum confinement
  - External resistance
  - Sub fin doping and gate all around devices
- Conclusion: How does the roadmap evolve?

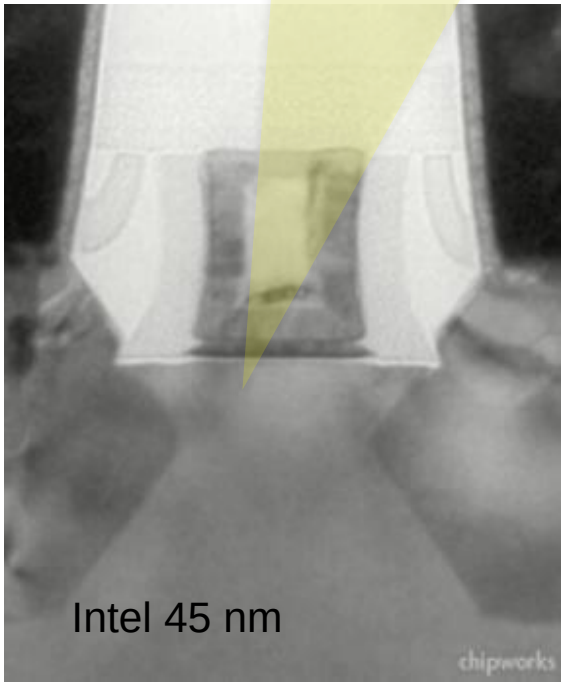
Drift Velocity and Velocity Saturation

Electric field = 0

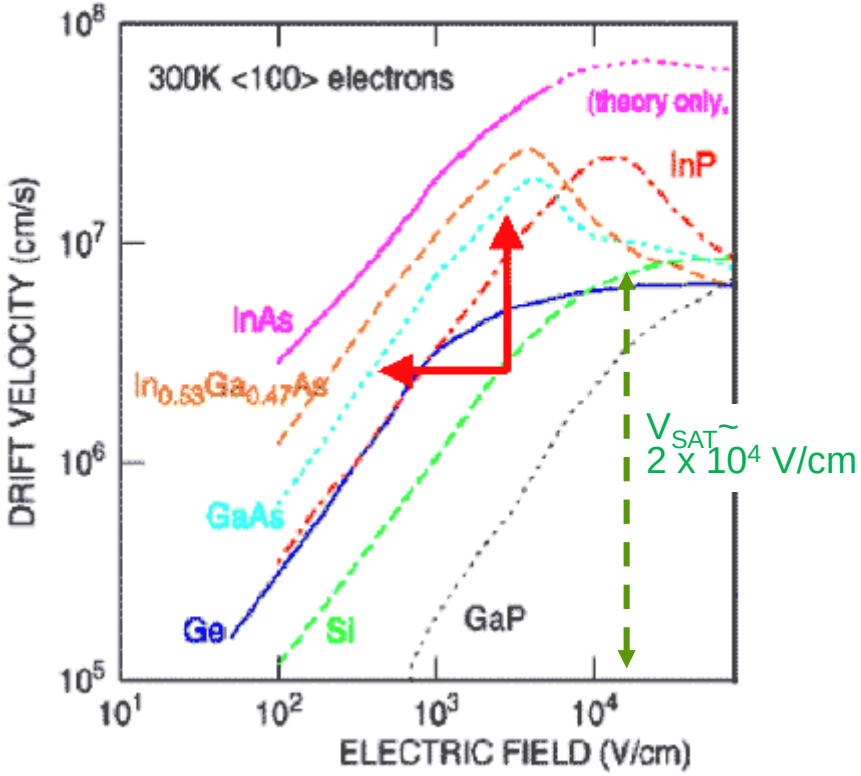


| Electric field | > 0





Drift velocity versus electric field for different semi-conductors. The red arrows show how drift velocity, i.e., mobility, can be improved at even lower electric fields by replacing silicon by other high-mobility semi-conductors.

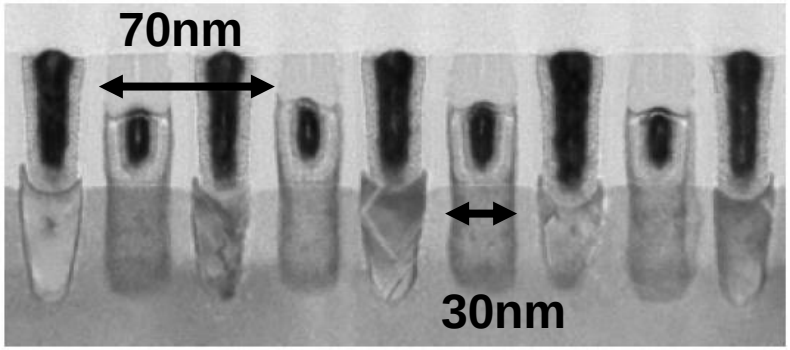


after M. Fischetti, 2001

Velocity Saturates: Nanoscale Device Over Most of the Channel

Intel 14 nm

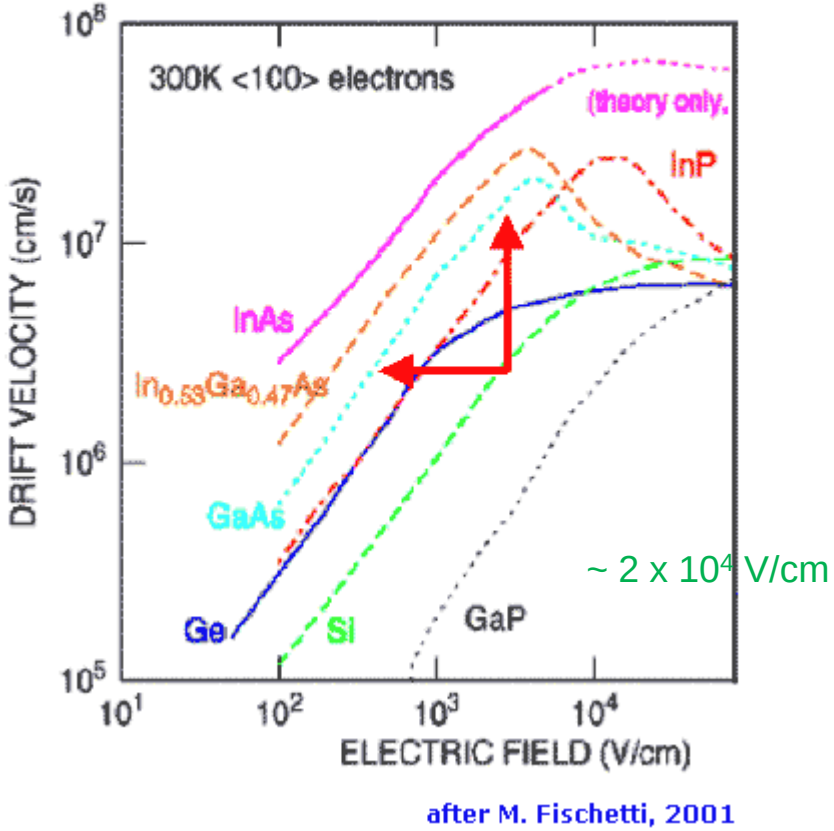
Source: Intel



$$L_{GATE} \approx 30nm \rightarrow L_{ELECTRICAL} \approx 20nm$$

$$\frac{V_{DS}}{L_E} \approx \frac{0.9V}{20nm} \approx 4.5x10^5 V/cm$$

Drift velocity versus electric field for different semiconductors. The red arrows show how drift velocity, i.e., mobility, can be improved at even lower electric fields by replacing silicon by other high-mobility semiconductors.

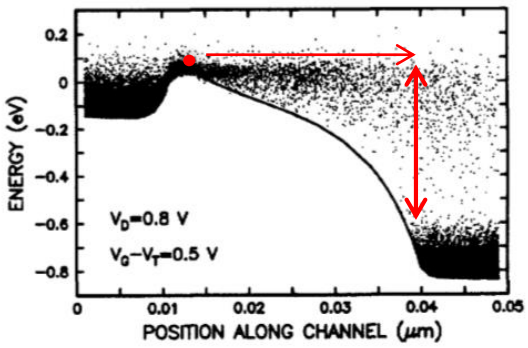


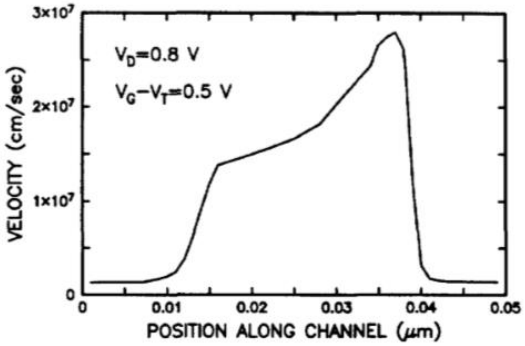
~ 2 x 10^4 V/cm

after M. Fischetti, 2001

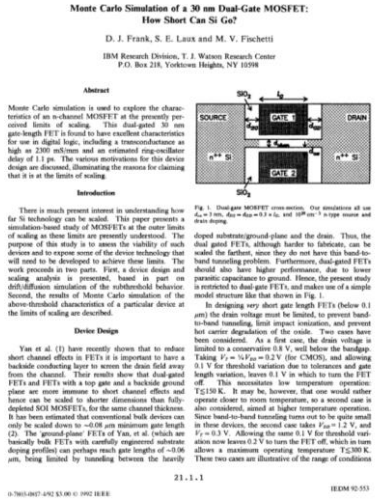
Key Concepts: Ballistic Transport

$\langle KE \rangle = \frac{1}{2} m^* v^2$





30nm Double Gate nFET



Monte Carlo Simulation of a 30 nm Dual-Gate MOSFET: How Short Can Si Go?

D. J. Frank, S. E. Laux and M. V. Fischetti

IBM Research Division, T. J. Watson Research Center
P.O. Box 218, Yorktown Heights, NY 10598

Abstract

Monte Carlo simulation is used to explore the characteristics of an n-channel MOSFET at the presently perceived limits of scaling. This dual-gated 30 nm gate-length FET is found to have excellent characteristics for use in digital logic, including a transconductance as high as 2300 mS/mm and an estimated ring-oscillator delay of 1.1 ps. The various motivations for this device design are discussed, illuminating the reasons for claiming that it is at the limits of scaling.

Introduction

There is much present interest in understanding how far Si technology can be scaled. This paper presents a simulation-based study of MOSFETs at the outer limits of scaling as these limits are presently understood. The purpose of this study is to assess the viability of such devices and to expose some of the device technology that will need to be developed to achieve these limits. The work proceeds in two parts. First, a device design and scaling analysis is presented, based in part on drift-diffusion simulation of the subthreshold behavior. Second, the results of Monte Carlo simulation of the above-threshold characteristics of a particular device at the limits of scaling are described.

Device Design

Yan et al. (1) have recently shown that to reduce short channel effects in FETs it is important to have a backside conducting layer to screen the drain field away from the channel. Their results show that dual-gated FETs and FETs with a top gate and a backside ground plane are more immune to short channel effects and hence can be scaled to shorter dimensions than fully-depleted SOI MOSFETs, for the same channel thickness. It has been estimated that conventional bulk devices can only be scaled down to ~0.08 μm minimum gate length (2). The 'ground-plane' FETs of Yan, et al. (which are basically bulk FETs with carefully engineered substrate doping profiles) can perhaps reach gate lengths of ~0.06 μm, being limited by tunneling between the heavily doped substrate/ground-plane and the drain. Thus, the dual gated FETs, although harder to fabricate, can be scaled the farthest, since they do not have this band-to-band tunneling problem. Furthermore, dual-gated FETs should also have higher performance, due to lower parasitic capacitance to ground. Hence, the present study is restricted to dual-gate FETs, and makes use of a simple model structure like that shown in Fig. 1.

In designing very short gate length FETs (below 0.1 μm) the drain voltage must be limited, to prevent band-to-band tunneling, limit impact ionization, and prevent hot carrier degradation of the oxide. Two cases have been considered. As a first case, the drain voltage is limited to a conservative 0.8 V, well below the bandgap. Taking $V_T \approx \frac{1}{4} V_{DD} = 0.2$ V (for CMOS), and allowing 0.1 V for threshold variation due to tolerances and gate length variation, leaves 0.1 V in which to turn the FET off. This necessitates low temperature operation: $T \lesssim 150$ K. It may be, however, that one would rather operate closer to room temperature, so a second case is also considered, aimed at higher temperature operation. Since band-to-band tunneling turns out to be quite small in these devices, the second case takes $V_{DD} = 1.2$ V, and $V_T = 0.3$ V. Allowing the same 0.1 V for threshold variation now leaves 0.2 V to turn the FET off, which in turn allows a maximum operating temperature $T \lesssim 300$ K. These two cases are illustrative of the range of conditions

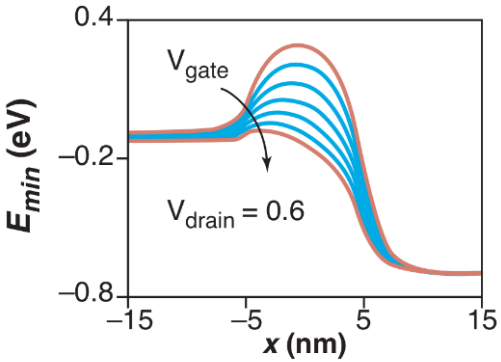
21.1.1

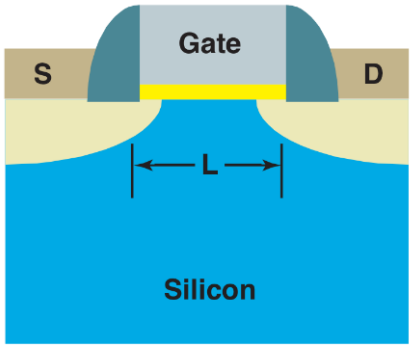
0-7803-0817-4/92 $3.00 © 1992 IEEE

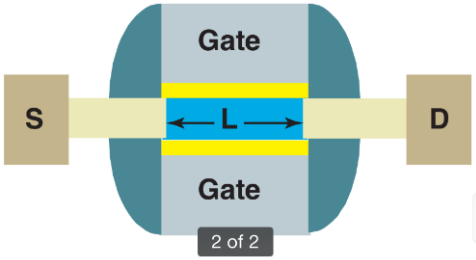
IEDM 92-553

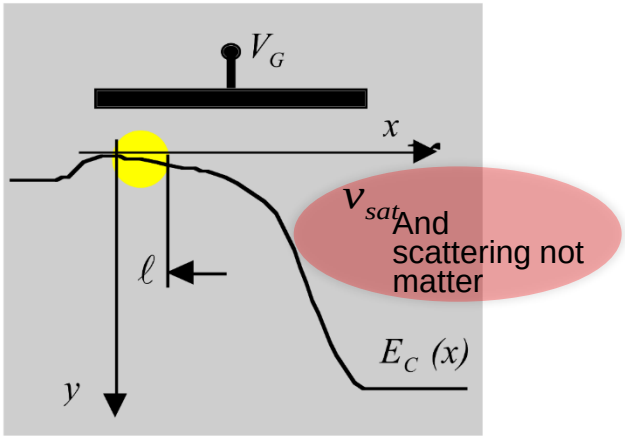
- **Source to drain transport "without" scattering**
- **L << λ —— Not right metric to near ballistic transport**
- **Scattering in high field region and velocity in high field does NOT matter**

Essential Physics of Nanoscale MOSFET









$$\frac{I_D}{W} = v \bullet Q_{inv} = v_T \bullet C_{OX}(V_{GS} - V_T)$$

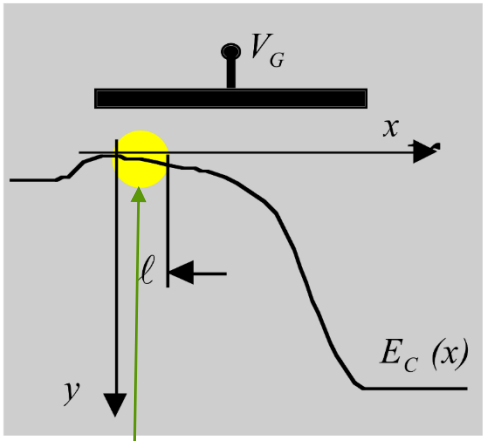
$$< v(o) > \approx (\frac{1-r}{1+r}) v_T$$

$$v_T \propto \sqrt{\frac{2k_B T_L}{\pi m_t^*}}$$

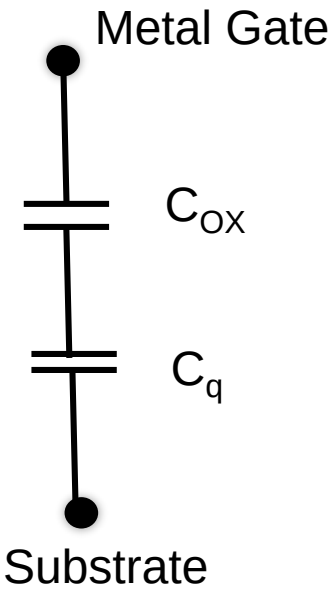
Source: Mark Lundstrum

- Carries injected into channel from “the source” / Gate largely controls source barrier
- Some fraction of carriers injected into channel backscatter and return to source
- Current increases as m^*_t decrease (why uniaxial stress)

Quantum Capacitance



$$\frac{I_D}{W} = v \bullet Q_{inv} = v_T \bullet C_{OX}(V_{GS} - V_T)$$



Modeled 2 capacitance in series

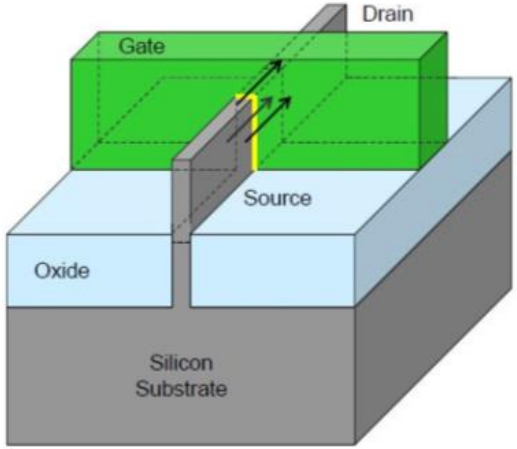
- Parallel-plate capacitor and low density of states in the semiconductor
- Capacitance is *not* given by the normal formula $\varepsilon_o\varepsilon_r/t_{ox}$ for parallel-plate capacitors
- Capacitance lowered due to another capacitor in series.
- Second capacitance (quantum capacitance) is important for low-density-of-states systems
- Lowers inversion charge for low density of state conduction band III-V semiconductors and 2D 2-dimensional electronic system (graphene et. al.)

Outline

- State-of-the-art for logic devices
- FinFET Basics
- Metrics for advance logic devices
- Deeper look at some advanced device concepts
  - Drive current and relation to mobility, velocity saturation, and density of states
  - **Transistor variation / random doping effect**
  - Strain I, Ge and III- channels
  - Quantum confinement
  - External resistance
  - Sub fin doping and gate all around devices
- Conclusion: How does the roadmap evolve?

FinFET: Many Additional Sources of Variation

- Multiple sources of variation (WID, WIW, WTW)
  - Fin height
    - Variation from : STI dep, STI CMP, wet etch
  - Fin width
  - Fin shape
  - Fin doping
  - Fin line edge roughness
  - Fin side wall plane
  - Gate LER (gate etch over fin topography)
  - Etched fin plane interface traps
  - Additional Epi variation (growth on etched plan)
  - SDE doping along fin height

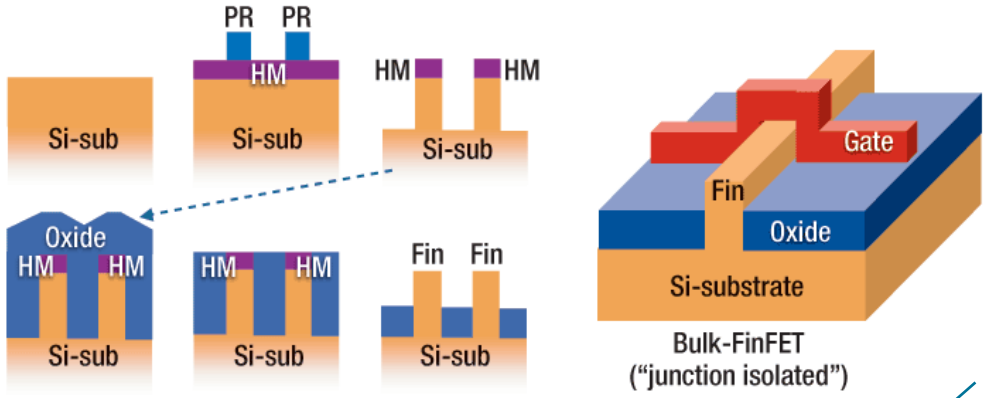


Source: Intel

Additional variation will be one of biggest challenges for SOC

Fin Height & Width Variation Large in Bulk FinFETs

- Additional sources of variation not present in bulk planar structures



Sources of variability	Nominal (nm)	3-sigma tolerance (current)	3-sigma tolerance (future)	
HM oxide	8	0.4	0.2	5% 3-sigma variation for oxide
HM nitride	70	7	3.5	10% 3-sigma variation for deposited nitride
Trench etch	170	8.5	4.25	5% 3-sigma from trench etch based on 32nm data
Oxide recess	100	5	2.5	Oxide dry/wet etch with no etch stop. 100nm oxide etchback for 70nm fin height assumed.
Pad oxide	2	0.1	0.05	
Well anneal	0	3	1.5	3-sigma variability in junction depth from angled implants
Total fin height variability (nm)		**12.5**	**6.2**	Root sum-square of all sources of variability
Total fin width variability (nm)		**2.5**	**1.2**	Assumption is that 20% of the vertical variability will translate to CD (Fin width) variability.

Table 2. Sources of variability for junction-isolated bulk FinFETS.

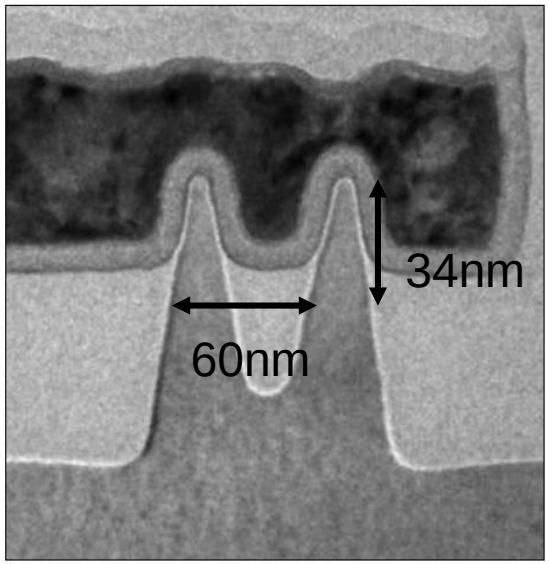
35% W variation

18% W variation

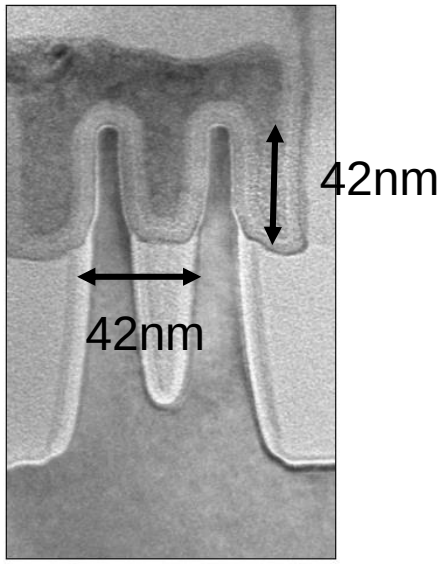
http://www.electroiq.com/articles/sst/print/volume-52/issue-11

Fin Scaling from 22nm to 14nm

Transistor Fin Improvement



22 nm 1st Generation
Tri-gate Transistor



14 nm 2nd Generation
Tri-gate Transistor

22/16/14nm FinFET Still Use Channel Dopants

$$N_{DOPANTS} = N_A WLW_{dep}$$

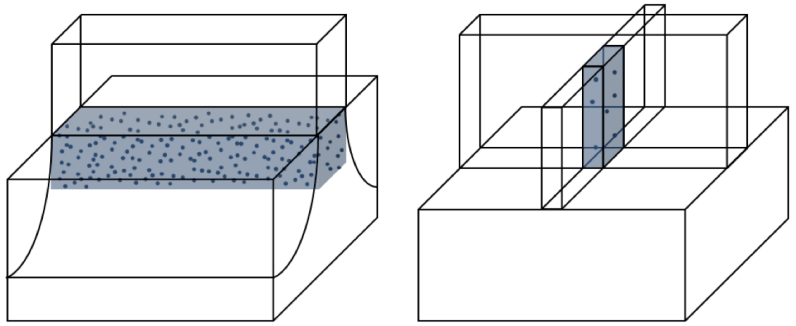
$$\sigma_{N_{DOPANTS}} = \sqrt{N_A WLW_{dep}}$$

$$\sigma_{Vt} = \frac{qT_{ox}}{\varepsilon_{ox}} \sqrt{\frac{N_{sub} W_{dep}}{3LW}}$$

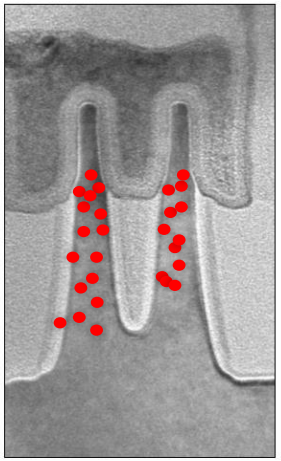
Reduced Channel Doping

Planar

Tri-Gate



Source: Intel/Bohr



- Doped glass used to dope sub-fin.
- Some will diffuse into bottom of FIN

Intel 22 and 14nm FinFET Data

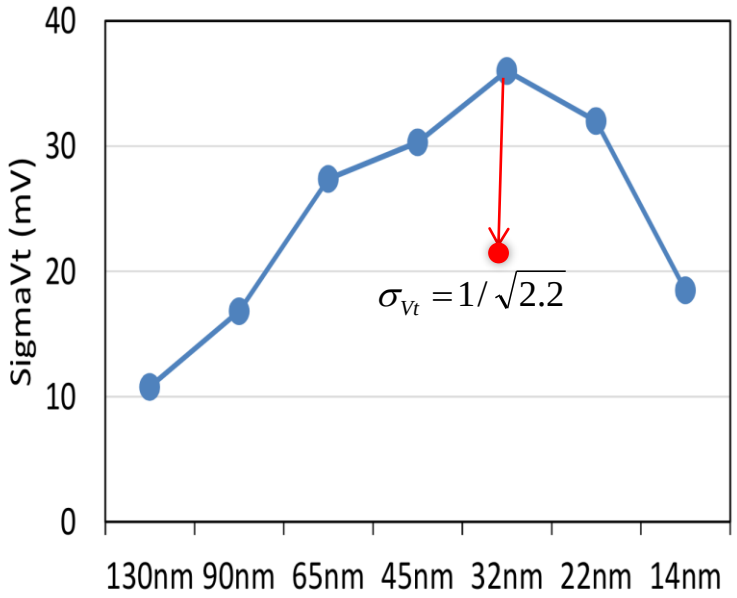


Figure 7: Random Varation Trend (σVt)

3 of 3

A 14nm Logic Technology Featuring 2nd-Generation FinFET Transistors, Air-Gapped Interconnects, Self-Aligned Double Patterning and a 0.0588μm² SRAM cell size

S. Natarajan, M. Agostinelli*, S. Akbar, M. Bost, A. Bowonder, V. Chikarmane, S. Chouksey, A. Dasgupta, K. Fischer, Q. Fu, T. Ghani, M. Giles**, S. Govindaraju, R. Grover, W. Han, D. Hanken, E. Haralson, M. Haran, M. Heckscher, R. Heussner, P. Jain, R. James, R. Jhaveri, I. Jin, H. Kam, E. Karl, C. Kenyon, M. Liu, Y. Luo, R. Mehandru**, S. Morarka**, L. Neiberg, P. Packan, A. Paliwal, C. Parker, P. Patel, R. Patel, C. Pelto, L. Pipes, P. Plekhanov, M. Prince, S. Rajamani, J. Sandford, B. Sell, S. Sivakumar, P. Smith, B. Song, K. Tone, T. Troeger, J. Wiedemer, M. Yang, K. Zhang

Logic Technology Development, * Quality and Reliability Engineering, ** DTS, Intel Corporation.
Phone: (503) 613-8029; Email: sanjay.natarajan@intel.com

ABSTRACT

A 14nm logic technology using 2nd-generation FinFET transistors with a novel subfin doping technique, self-aligned double patterning (SADP) for critical patterning layers, and air-gapped interconnects at performance-critical layers is described. The transistors feature rectangular fins with 8nm fin width and 42nm fin height, 4th generation high-k metal gate, and 6th-generation strained silicon, resulting in the highest drive currents yet reported for 14nm technology. This technology is in high-volume manufacturing.

INTRODUCTION

Relentless focus on transistor scaling and Moore's Law has led to ever-higher transistor performance and density, translating into tremendous increases in microprocessor functionality and performance. Traditional scaling had reached limitations of electrostatics and short-channel effects, threatening the continuance of Moore's Law. 22nm tri-gate transistors[1] broke through some of these scaling barriers. This paper reports on a 14nm process technology, including a 2nd generation FinFET architecture, which provides industry-leading transistor performance and density.

KEY DESIGN RULES & TECHNOLOGY FEATURES

Table 1 summarizes key design rules and scale factors for the 14nm node compared to 22nm node[1]. Fin pitch, a key measure of transistor density for FinFETs, is scaled to 42nm, maintaining 0.7x scaling trend from 22nm[1]. Contacted gate pitch is scaled to 70nm. Interconnect pitches are as low as 52nm and scaled between 0.65x and 0.78x. SADP with 193nm immersion lithography is used at critical patterning layers to enable aggressive design rule scaling. Fig. 1 shows the scaling trend of contacted gate pitch multiplied by metal pitch (as a proxy for density) for the past four Intel generations as well as published industry data. The Intel 14nm process continues to maintain the historical trend of density improvement per generation.

TRANSISTOR PERFORMANCE AND VARIATION

Lgate scaling from 26nm in the 22nm node (from conference presentation from [1]) to 20nm in the 14nm node is enabled by fin profile optimization and a novel subfin doping technique. Subfin doping of high performance transistors is achieved through solid-source doping to enable better optimization of punch-through stopper dopants. Fig. 2 shows transistor fin-cut and gate-cut images. NMOS and PMOS Idsat/Ioff and Idlin/Ioff curves are shown in Figs. 3 and 4. At 0.7V Vdd, 10nA/um Ioff, 42nm fin pitch and 70nm contacted gate pitch, saturated drive currents are 1.04mA/um (all drive currents are per-micron of layout width) for both NMOS and PMOS. Idsat is improved 15% for NMOS and 41% for PMOS over 22nm[1], and these are the best drive currents reported to-date for 14nm technology. NMOS and PMOS linear drive currents are 0.237mA/um and 0.203mA/um, respectively, at 10nA/um Ioff, Vgs=0.7V, and Vds=50mV. These represent an improvement of 30% for NMOS and 38% for PMOS over 22nm[1]. Transistor I-V and sub-threshold characteristics are shown in Figs. 5 and 6. Sub-threshold slopes are maintained at ~65mV/decade. DIBL is ~60mV/V and ~75 mV/V for NMOS and PMOS, respectively. Random variation data for minimum-sized devices for multiple Intel process technology generations is shown in Fig. 7. The trend in increasing σVt due to Z scaling for planar MOSFETs was reversed in 22nm. With optimizations in fin profile and doping, σVt was reduced in 14nm by a factor of nearly 2x, enabling significant improvement in product Vmin.

RELIABILITY

Optimization of the high-k + metal-gate stack yields excellent reliability characteristics. Fig. 8 shows PMOS and NMOS TDDB, respectively. Both show a clear improvement relative to 22nm. Fig. 9 shows superior aging compared to 22nm as measured by reduced Vccmin degradation under stress.

INTERCONNECTS

10 layers of the 13-layer Cu interconnect stack are shown in Fig. 10. Low-k CDO dielectrics are used on 8 layers. Lower-layer metal pitches are scaled aggressively relative to [1], by as much as 0.65x. Fig. 11 shows two layers of air-gapped interconnect, used to provide improved capacitance at performance-critical layers. Air gaps are used at 80nm and 160nm minimum pitch layers and provide a 17% improvement in capacitance. A thick top metal is used for improved on-die power distribution.

SRAM, PRODUCT AND YIELD

The 14nm test chip includes a 140Mbit SRAM featuring a 0.0588μm² bitcell (Fig. 12) with an array density of 11.6Mb/mm². This process is now in high-volume manufacturing, producing a family of processors using the Broadwell microarchitecture (Fig. 13). Yield in production is in a healthy range for high-volume manufacturing.

CONCLUSIONS

We have demonstrated an industry-leading 14nm CMOS technology with excellent transistor and interconnect performance and aggressive design rule scaling. The process features 2nd generation FinFETs with optimized fin profiles and a novel sub-fin doping technique, 52nm minimum pitch interconnects formed with SADP, air-gap interconnects at performance-critical layers, and aggressive design rule scaling from the 22nm generation to provide a true 14nm technology. We have shown a high-performance, high-density SRAM featuring 0.0588μm² cell size fabricated using all 14nm process features. This process is in high-volume manufacturing.

REFERENCES

1. C. Auth, et al., *Symp. VLSI Tech. Dig.*, p. 131, 2012.
2. K-L Cheng, et al., *IEDM Tech. Dig.*, p. 243, 2007.
3. F. Arnaud, et al., *IEDM Tech. Dig.*, p. 651, 2009.
4. H. Shang, et al., *Symp. VLSI Tech Dig.*, p.129, 2012.
5. S. Wu, et al., *IEDM Tech Dig.*, p. 224, 2013.
6. K-I Seo, et al., *Symp. VLSI Tech. Dig.*, p. 14, 2014.

- **Note data is not normalized to W as for Planar**
- **Large finFET W is a REAL advantage for SRAM and is responsible for significant part of σ_{VT}**

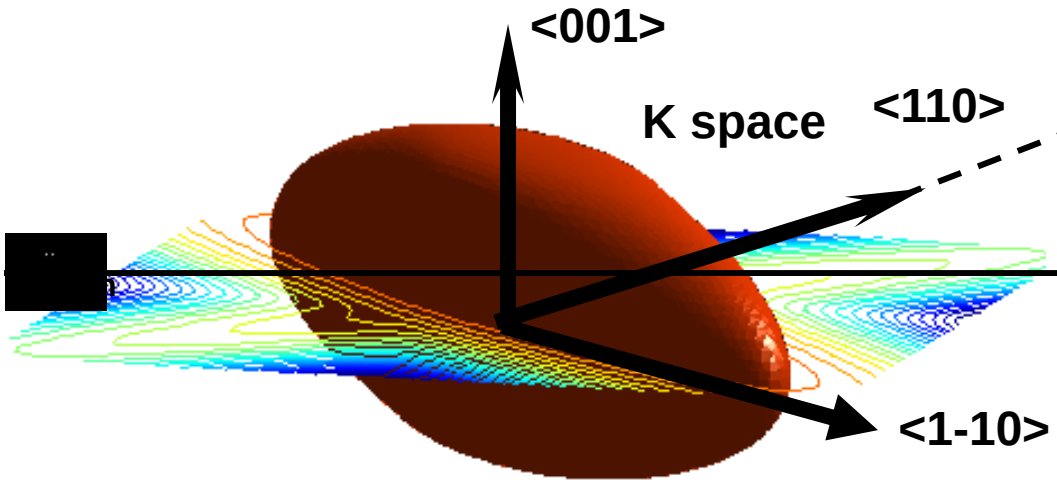
Outline

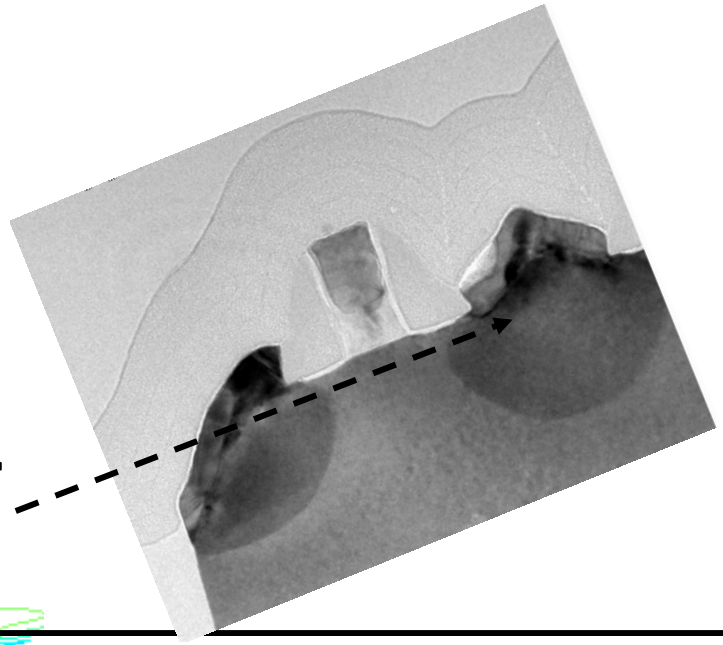
- State-of-the-art for logic devices
- FinFET Basics
- Metrics for advance logic devices
- Deeper look at some advanced device concepts
  - Drive current and relation to mobility, velocity saturation, and density of states
  - Transistor variation / random doping effect
  - **Strain I, Ge and III- channels**
  - Quantum confinement
  - External resistance
  - Sub fin doping and gate all around devices
- Conclusion: How does the roadmap evolve?

What is the Desired Band Structure

$$\mu = \frac{q\tau}{m^*}$$

Out of Plane: Large m* for quantum confinement to be additive with strain

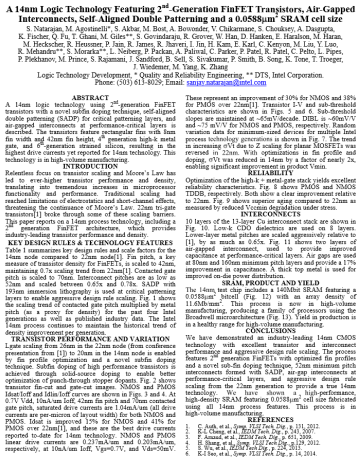




Channel Direction: Small m* for high mobility

In-plane $\perp$ to Channel: Large m* for high density-of-states

What "Velocity" Sets I_{ON} and Why is N/P Ratio ~1



A 14nm Logic Technology Featuring 2nd-Generation FinFET Transistors, Air-Gapped Interconnects, Self-Aligned Double Patterning and a 0.0588μm² SRAM cell size

S. Natarajan, M. Agostinelli*, S. Akbar, M. Bost, A. Bowonder, V. Chikarmane, S. Chouksey, A. Dasgupta, K. Fischer, Q. Fu, T. Ghani, M. Giles**, S. Govindaraju, R. Grover, W. Han, D. Hanken, E. Haralson, M. Haran, M. Heckscher, R. Heussner, P. Jain, R. James, R. Jhaveri, I. Jin, H. Kam, E. Karl, C. Kenyon, M. Liu, Y. Luo, R. Mehandru**, S. Morarka**, L. Neiberg, P. Packan, A. Paliwal, C. Parker, P. Patel, R. Patel, C. Pelto, L. Pipes, P. Plekhanov, M. Prince, S. Rajamani, J. Sandford, B. Sell, S. Sivakumar, P. Smith, B. Song, K. Tone, T. Troeger, J. Wiedemer, M. Yang, K. Zhang

Logic Technology Development, * Quality and Reliability Engineering, ** DTS, Intel Corporation.
Phone: (503) 613-8029; Email: sanjay.natarajan@intel.com

ABSTRACT

A 14nm logic technology using 2nd-generation FinFET transistors with a novel sub5ns doping technique, self-aligned double patterning (SADP) for critical patterning layers, and air-gapped interconnects at performance-critical layers is described. The transistors feature rectangular fins with 8nm fin width and 42nm fin height, 4th generation high-k metal gate, and 6th-generation strained silicon, resulting in the highest drive currents yet reported for 14nm technology. This technology is in high-volume manufacturing.

INTRODUCTION

Relentless focus on transistor scaling and Moore's Law has led to ever-higher transistor performance and density, translating into tremendous increases in microprocessor functionality and performance. Traditional scaling had reached limitations of electrostatics and short-channel effects, threatening the continuance of Moore's Law. 22nm tri-gate transistors[1] broke through some of these scaling barriers. This paper reports on a 14nm process technology, including a 2nd generation FinFET architecture, which provides industry-leading transistor performance and density.

KEY DESIGN RULES & TECHNOLOGY FEATURES

Table I summarizes key design rules and scale factors for the 14nm node compared to 22nm node[1]. Fin pitch, a key measure of transistor density for FinFETs, is scaled to 42nm, maintaining 0.7x scaling trend from 22nm[1]. Contacted gate pitch is scaled to 70nm. Interconnect pitches are as low as 52nm and scaled between 0.65x and 0.7lx. SADP with 193nm immersion lithography is used at critical patterning layers to enable aggressive design rule scaling. Fig 1 shows the scaling trend of contacted gate pitch multiplied by metal pitch (as a proxy for density) for the past three Intel generations as well as published industry data. The Intel 14nm process continues to track the historical trend of ~0.5x process-to-process.

TRANSISTOR PERFORMANCE AND VARIATION

Lgate scaling from 26nm in the 22nm node (from conference presentation from [1]) to 20nm in the 14nm node is enabled by fin profile optimization and a novel sub5ns doping technique. Sub5ns doping of high performance transistors is achieved through solid-source doping to enable better optimization of punch-through stopper dopants. Fig. 2 shows transistor fin-cut and gate-cut images. NMOS and PMOS Idsat/Ioff and Idlin/Ioff curves are shown in Figs. 3 and 4. At 0.7V Vdd, 10nA/um Ioff, 42nm fin pitch and 70nm contacted gate pitch, saturated drive currents are 1.04mA/um (all drive currents are per-micron of layout width) for both NMOS and PMOS. Idsat is improved 15% for NMOS and 41% for PMOS over 22nm[1], and these are the best drive currents reported to-date for 14nm technology. NMOS and PMOS linear drive currents are 0.237mA/um and 0.203mA/um, respectively, at 10nA/um Ioff, Vgs=0.7V, and Vds=50mV. These represent an improvement of 30% for NMOS and 38% for PMOS over 22nm[1]. Transistor I-V and sub-threshold characteristics are shown in Figs. 5 and 6. Sub-threshold slopes are maintained at ~67mV/decade. DIBL is ~40mV/V and ~73 mV/V for NMOS and PMOS, respectively. Random variation data for minimum-sized devices for multiple lateral process technology generations is shown in Fig. 7. The trend in increasing σVt due to Z scaling for planar MOSFETs was reversed in 22nm. With optimizations in fin profile and doping, σVt was reduced in 14nm by a factor of nearly 2x, enabling significant improvement in product Vmin.

RELIABILITY

Optimization of the high-k + metal-gate stack yields excellent reliability characteristics. Fig. 8 shows PMOS and NMOS TDDB, respectively. Both show a clear improvement relative to 22nm. Fig. 9 shows superior aging compared to 22nm as measured by reduced Vccmin degradation under stress.

INTERCONNECTS

10 layers of the 13-layer Cu interconnect stack are shown in Fig. 10. Low-k CDO dielectrics are used on 8 layers. Lower-layer metal pitches are scaled aggressively relative to [1], by as much as 0.65x. Fig. 11 shows two layers of self-aligned double patterning used to provide improved resistance at performance-critical layers. Air gaps are used at 80nm and 160nm minimum pitch layers and provide a 17% improvement in capacitance. A thick top metal is used for improved on-die power distribution.

SRAM, PRODUCT AND YIELD

The 14nm test chip includes a 140Mbit SRAM featuring a 0.0588μm² 3stroll (Fig. 12) with an array density of 11.4Mb/mm². This process is now in high-volume manufacturing, producing a family of processors using the Broadwell microarchitecture (Fig. 13). Yield in production is in a healthy range for high-volume manufacturing.

CONCLUSIONS

We have demonstrated an industry-leading 14nm CMOS technology with excellent transistor and interconnect performance and aggressive design rule scaling. The process features 2nd generation FinFETs with optimized fin profiles and a novel sub-5ns doping technique, 52nm minimum pitch interconnects formed with SADP, air-gap interconnects at performance-critical layers, and aggressive design rule scaling from the 22nm generation to provide a true 14nm technology. We have shown a high-performance, high-density SRAM featuring 0.0588μm² cell size fabricated using all 14nm process features. This process is in high-volume manufacturing.

REFERENCES

1. C. Auth, et al., Symp. VLSI Tech. Dig., p.131, 2012.
2. K.-J. Chang, et al., IEDM Tech. Dig., p. 27.1.1, 2007.
3. F. Arnaud, et al., Symp. VLSI Tech. Dig., p. 197, 2009.
4. H. Shang, et al., Symp. VLSI Tech. Dig., p.129, 2012.
5. S. Wu, et al., IEDM Tech. Dig., p. 256, 2011.
6. K-I Seo, et al., Symp. VLSI Tech. Dig., p. 14, 2014.

SiGe Stress
$m^*_t \sim m^*_{lh} \sim 0.15 m_o$

$$\frac{I_D}{W} = v \bullet Q_{inv} = v \bullet C_{OX}(V_{GS} - V_T)$$

$$v = \mu E \; ?$$

$$v = v_{SAT} \; ?$$

$$v = v_T \; ? \qquad v_T \propto \sqrt{\frac{2k_B T_L}{\pi m_t^*}}$$

N/P ratio = I^N_{ON}/I^P_{ON} ~ 1.1 !

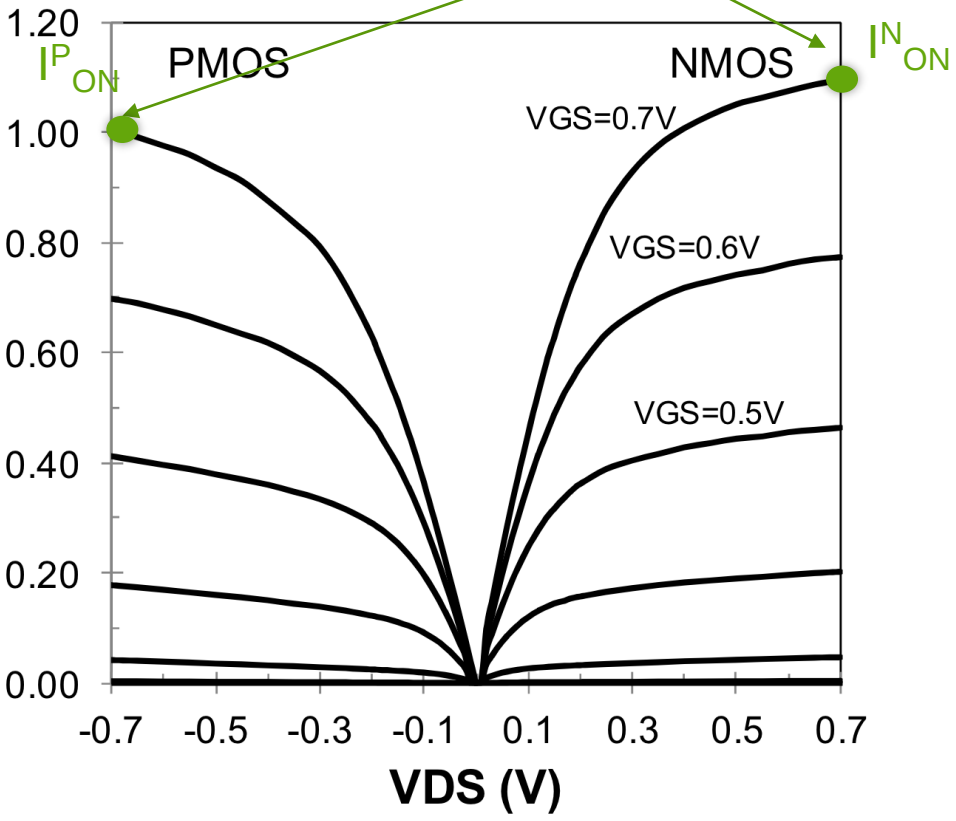
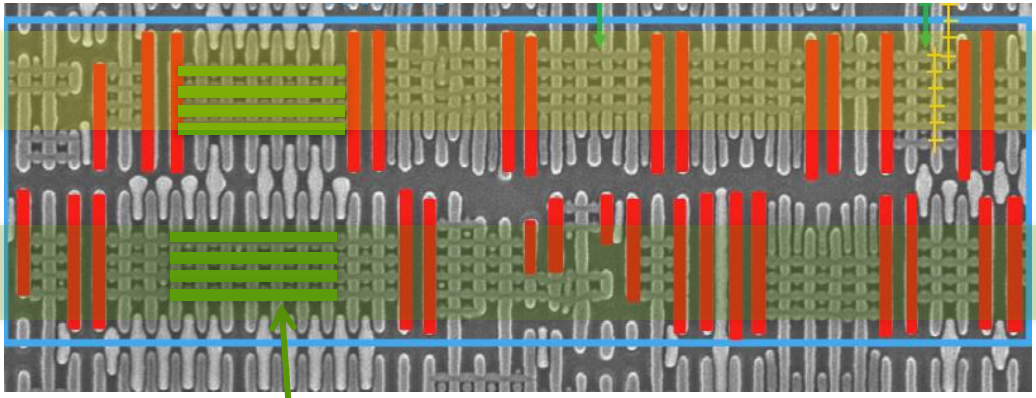


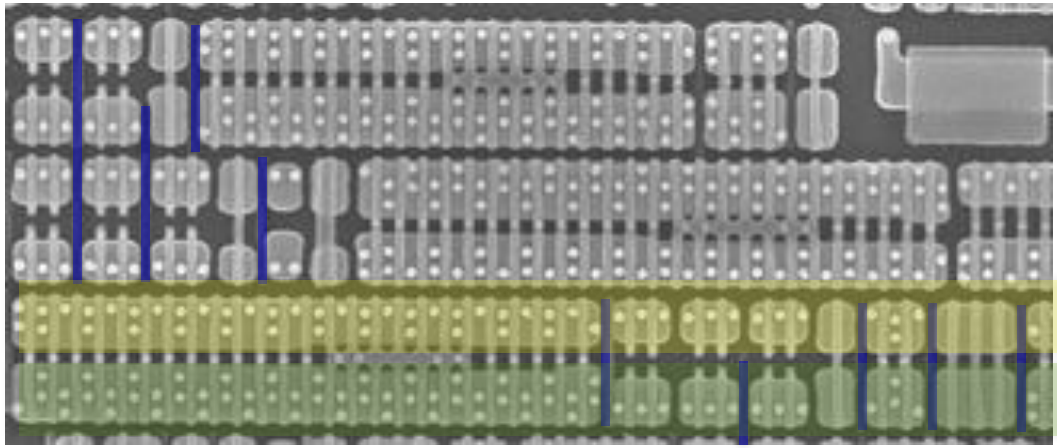
Figure 5: Transistor I-V Curves

Intel 14nm Transistor

Layout of Real Chips with SiGe Stressors



Intel 22nm

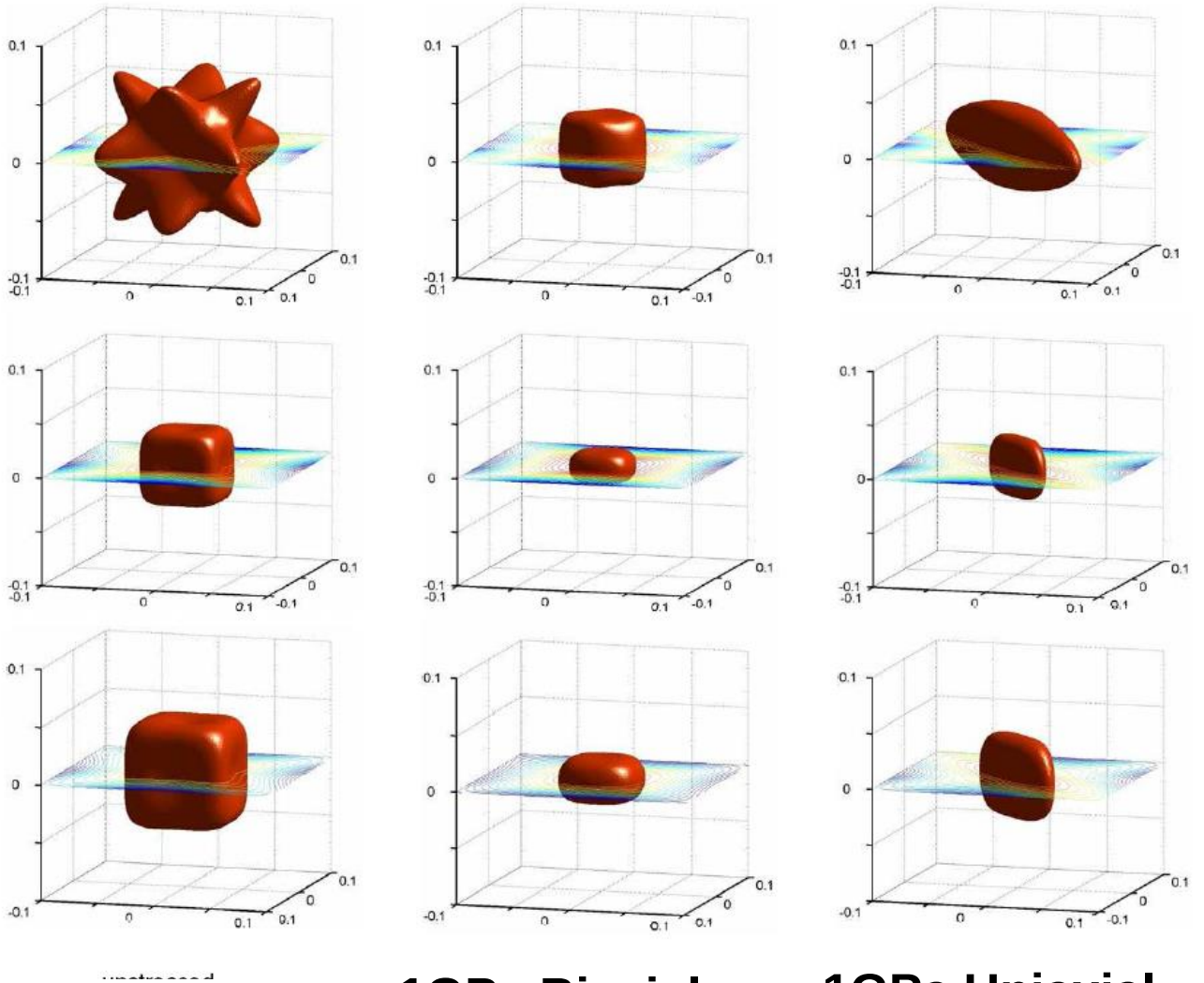


Qualcomm 28nm

Source : Chipworks

- SiGe stress allows for N/P ~1
- SiGe stress reduced pmos W by factor of 2
- Improves layout density and lowers C_{GATE}

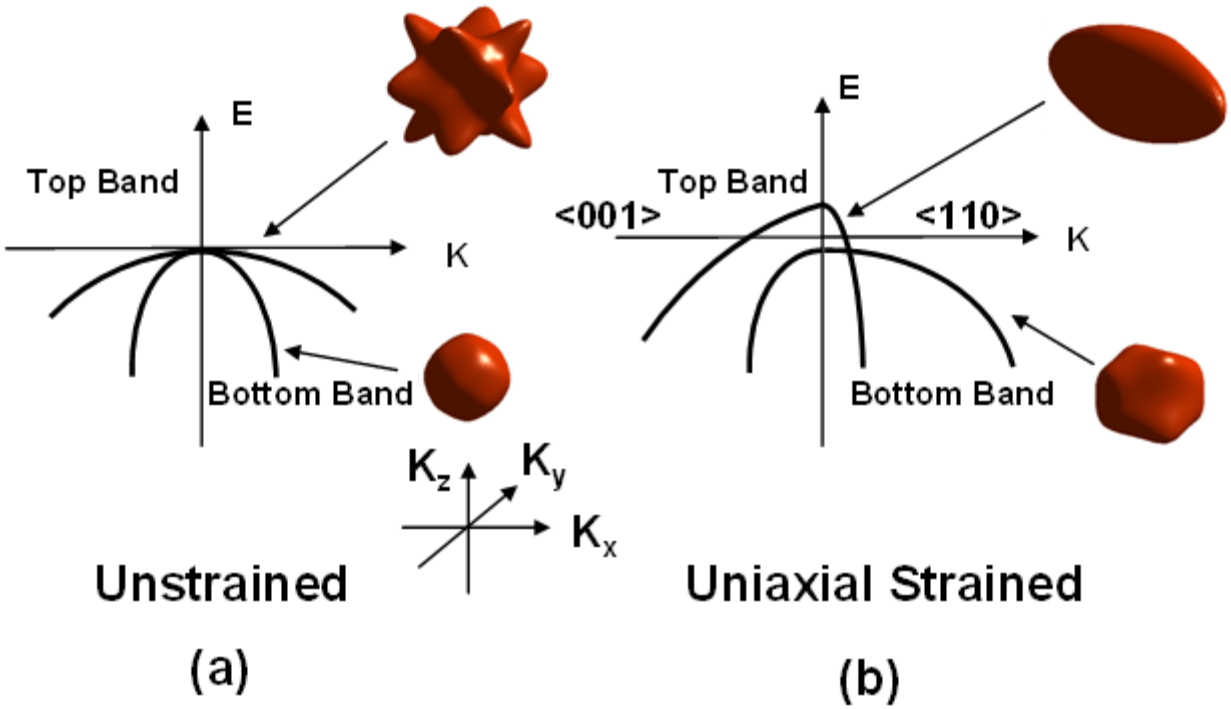
Same Physics for IV, V Materials



Strain Enhancement: Simple Estimate

$$\propto \frac{m_{hh}}{m_{lh}}$$

Conductivity mass ratio



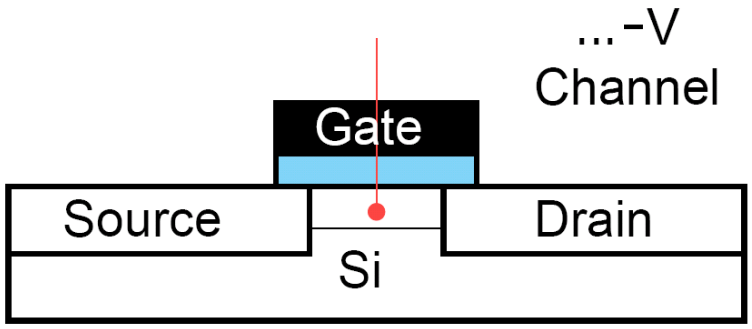
III-V Channel: Inversion Mode Device

TABLE III. Band and transport parameters for Si, Ge, and some III–V semiconductors.

	E_g (eV)	m_n	μ_n (cm^2/V s)	m_{hh}/m_{lh}	μ_h (cm^2/V s)	g_{mn} (Norm.)	$\Delta E_{\Gamma L}$ [c] (eV)	$m_l/m_t(L)$ [c]
Si	1.12	0.92/0.19	1450	0.53/0.15	500	1	⋯	⋯
Ge	0.67	1.59/0.082	3900	0.33/0.043	2270	0.92	⋯	⋯
InSb	0.17	0.014	7.7×10^4	0.45/0.016	850	1.9	0.51	1.56/0.094
InAs	0.35	0.024	$2-3.3\times10^4$	0.41/0.026	100–450	1.29	0.72	1.56/0.094
GaSb	0.73	0.041	3750	0.40/0.05	680	0.28	0.084	0.95/0.11
InP	1.34	0.08	5370	0.6/0.089	150	0.77	0.59	1.9/0.15
GaAs	1.42	0.063	9200	0.5/0.076	400	1.03	0.29	1.9/0.075
Strained Si	1.08	⋯	2900[a]	⋯	2200[b]	2	⋯	⋯

[a]See Ref. 26.
[b]See Ref. 99.
[c]These two columns list the L valley data for selected III–V semiconductors. The other data in the table are taken from Refs. 134 and 135.

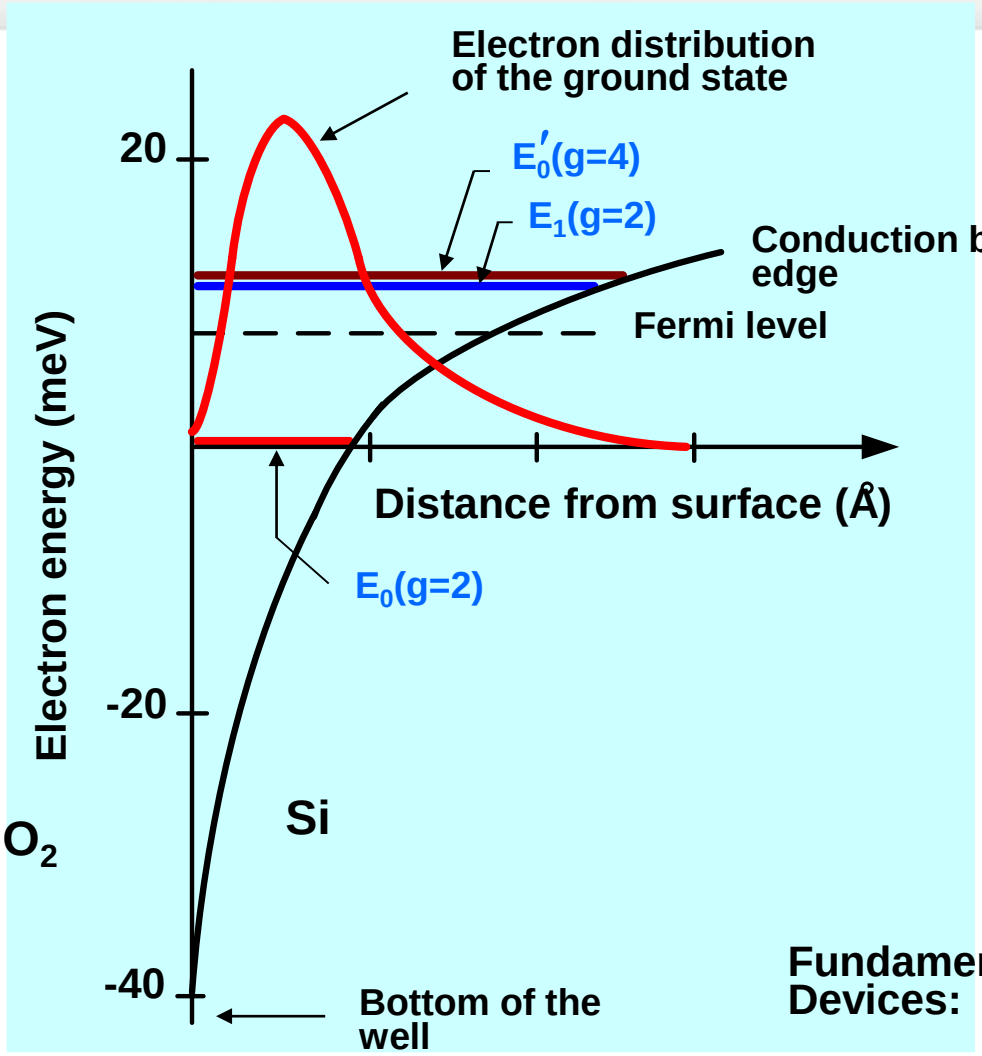
- **Ge and III-V channel pFET will benefit greatly by strain and will be needed for performance over Si**



Outline

- State-of-the-art for logic devices
- FinFET Basics
- Metrics for advance logic devices
- Deeper look at some advanced device concepts
  - Drive current and relation to mobility, velocity saturation, and density of states
  - Transistor variation / random doping effect
  - Strain I, Ge and III- channels
  - **Quantum confinement**
  - External resistance
  - Sub fin doping and gate all around devices
- Conclusion: How does the roadmap evolve?

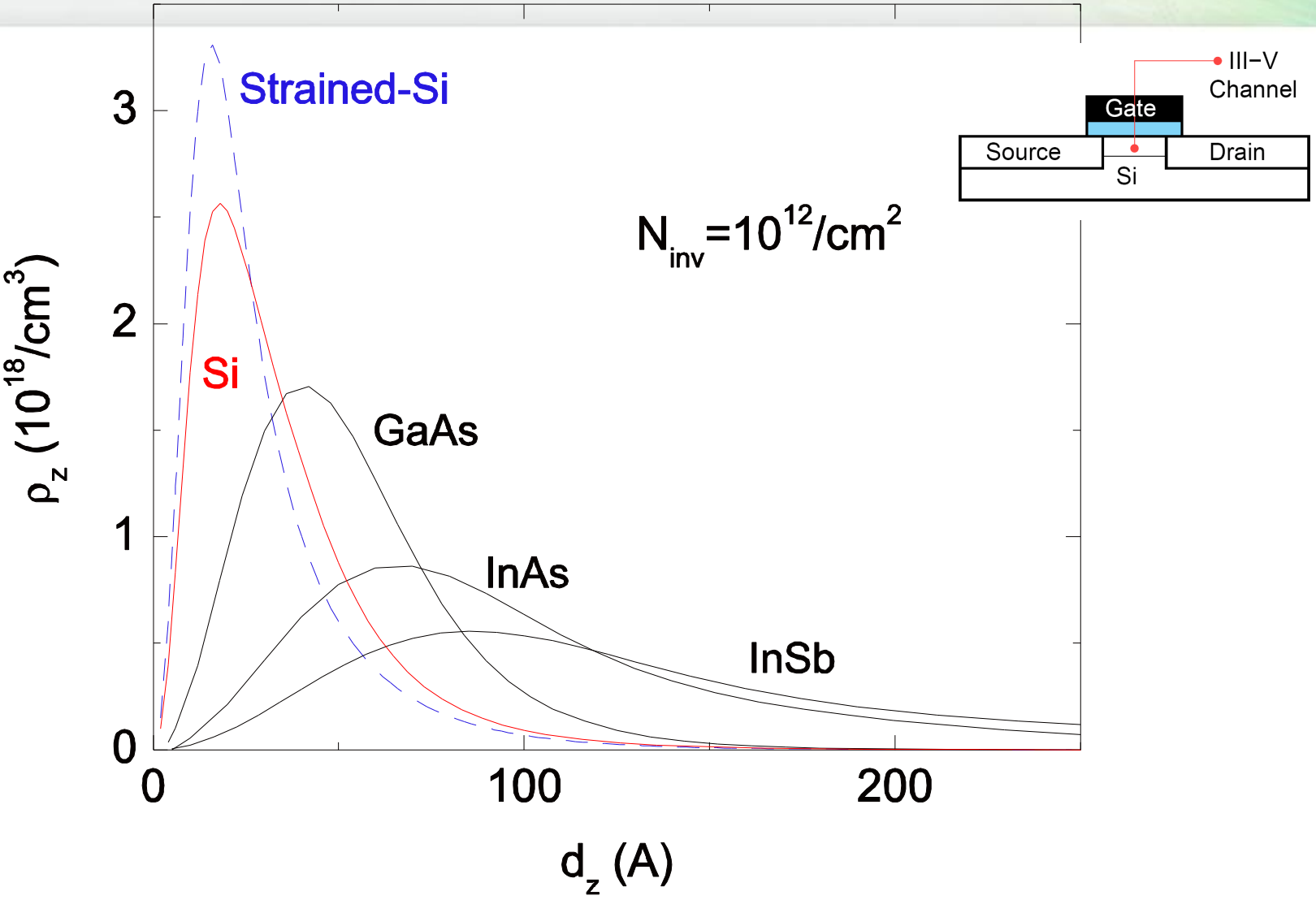
MOSFET inversion layer Confinement



Fundamentals of Modern VLSI Devices: Taur, Ning

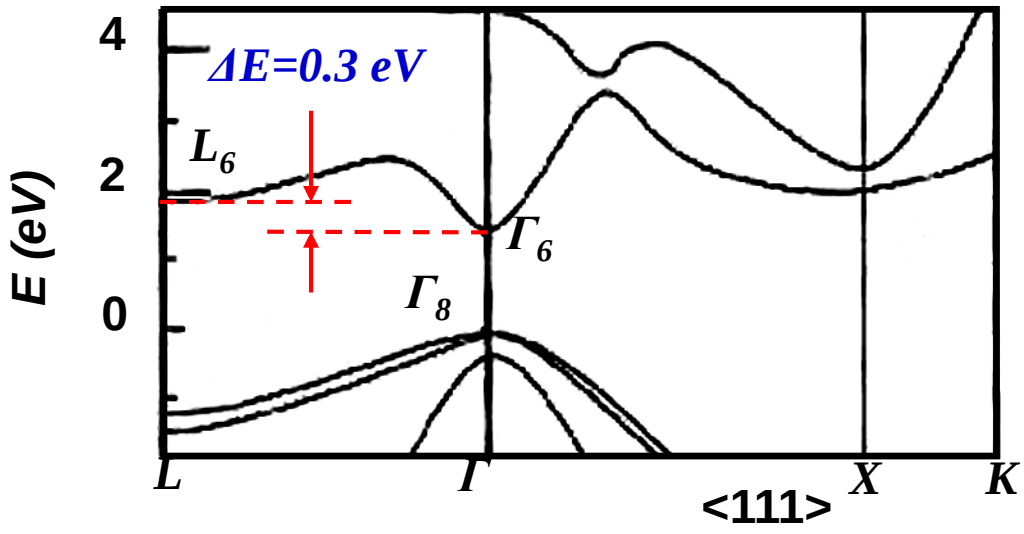
- Conduction band forms a potential well and quantum confinement occurs
- Carriers in bound states cannot propagate to infinity
- Bulk confinement potential

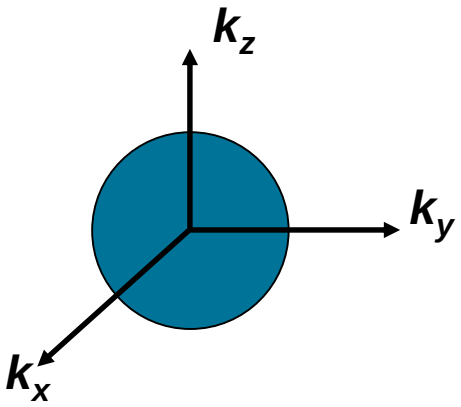
Location of Inversion Charge (NFET)



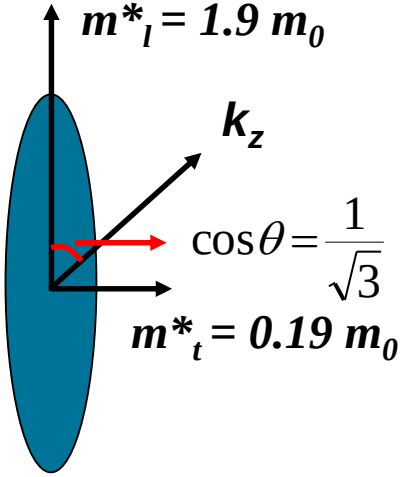
- **For Lighter confinement mass, inversion charge centroid further into semiconductor**

Case Study of GaAs



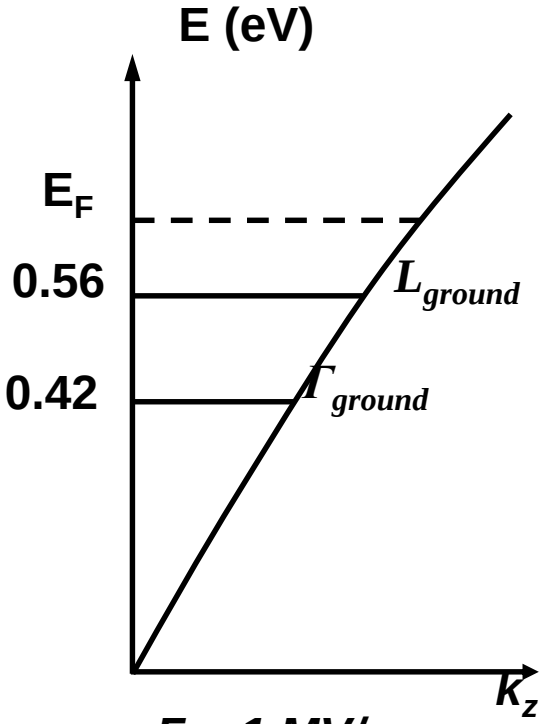


Γ_6 : spherical, $m^* = 0.067\ m_0$



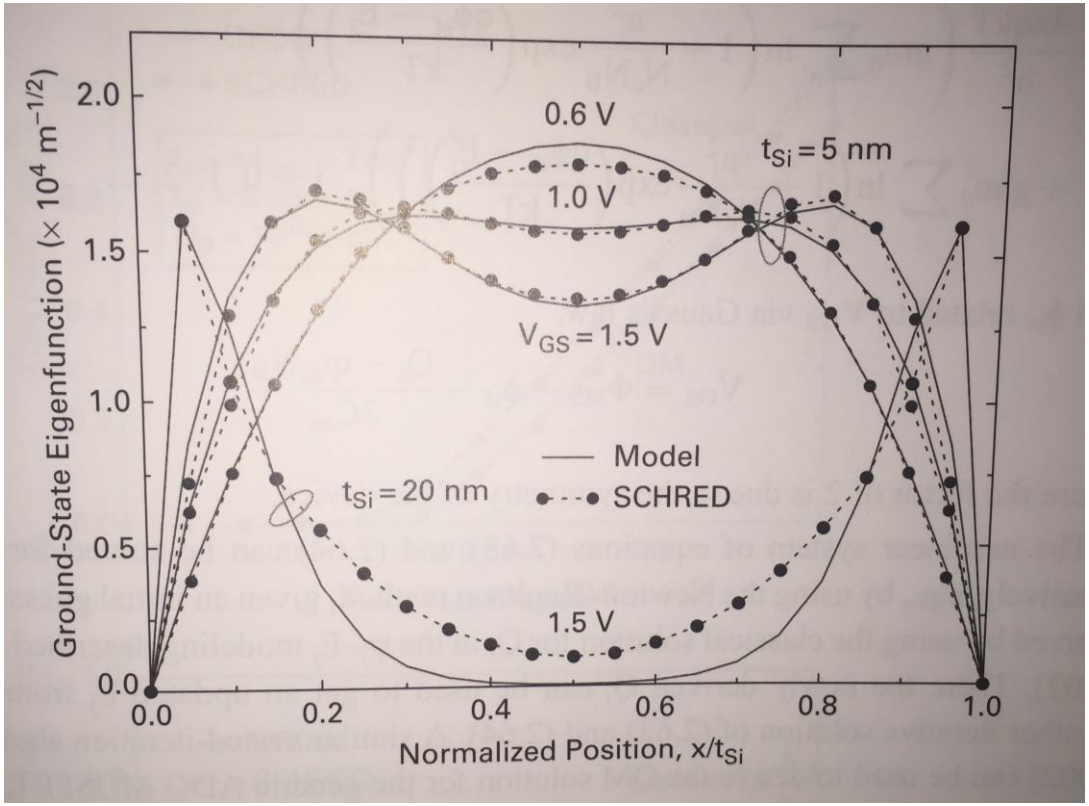
L_6 : ellipsoidal, $m_z = 0.27\ m_0$

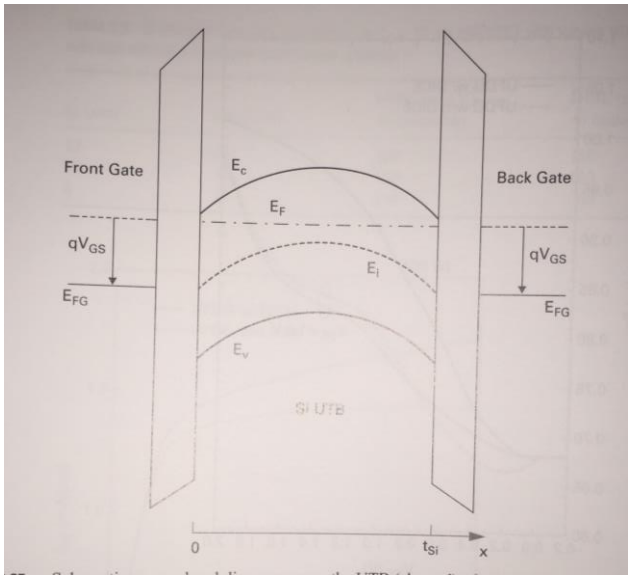
GaAs under (001) electric confinement



F = 1 MV/cm

Quantum Confinement FinFET





Source J.G. Fossum and V. P. Trivedi

Real Issue for 10/7nm node

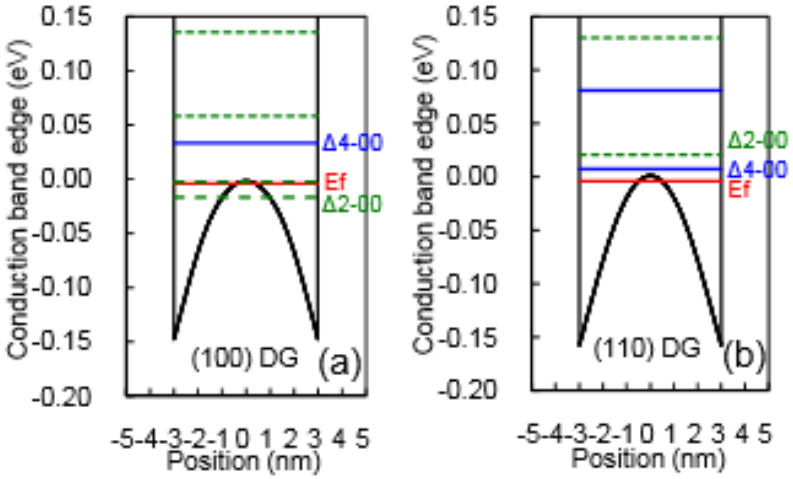


Fig.6 Sub-band distributions for **(a)** (100), and **(b)** (110) DG nMOSFETs with T_{Si} = 6 nm when $N_{inv} \approx 10^{13}$ cm^{-2}.

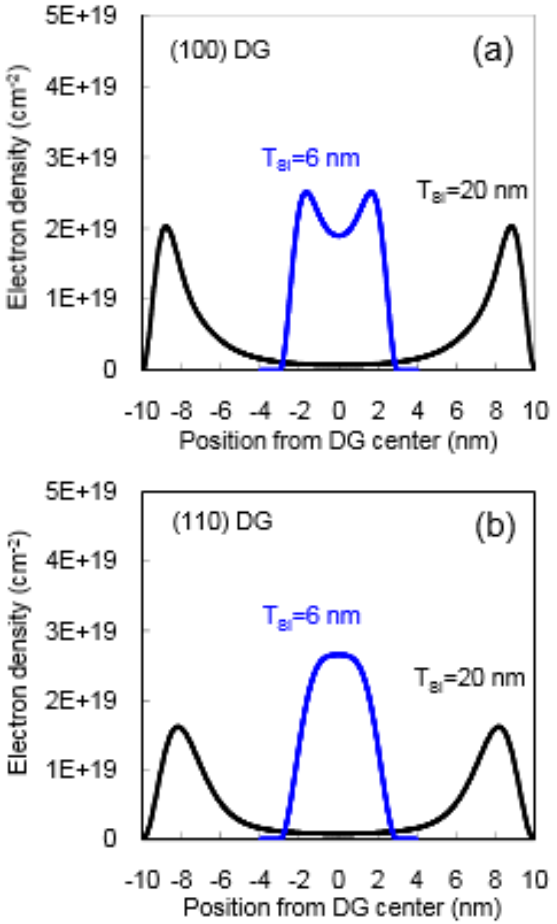


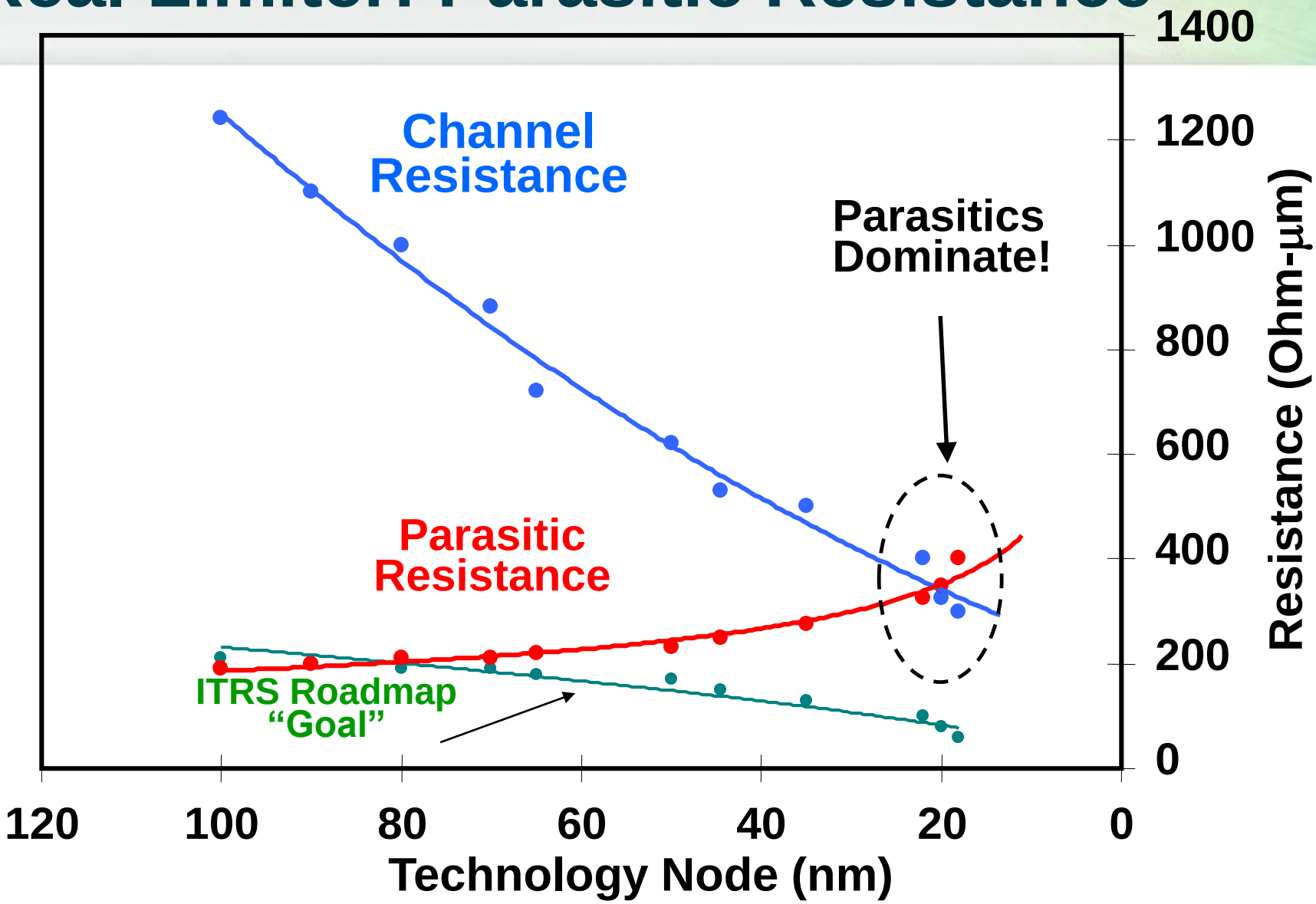
Fig.7 The charge distributions for **(a)** (100), and **(b)** (110) DG nMOSFETs with T_{Si} = 20 (black line) and 6 (blue line) nm. All charge distributions are plotted for $N_{inv} \approx 10^{13}$ cm^{-2}.

Source: Anson C-C Wang, Edward Chen, Tzer-Min Shen, Jeff Wu, and Carlos H. Diaz. TSMC

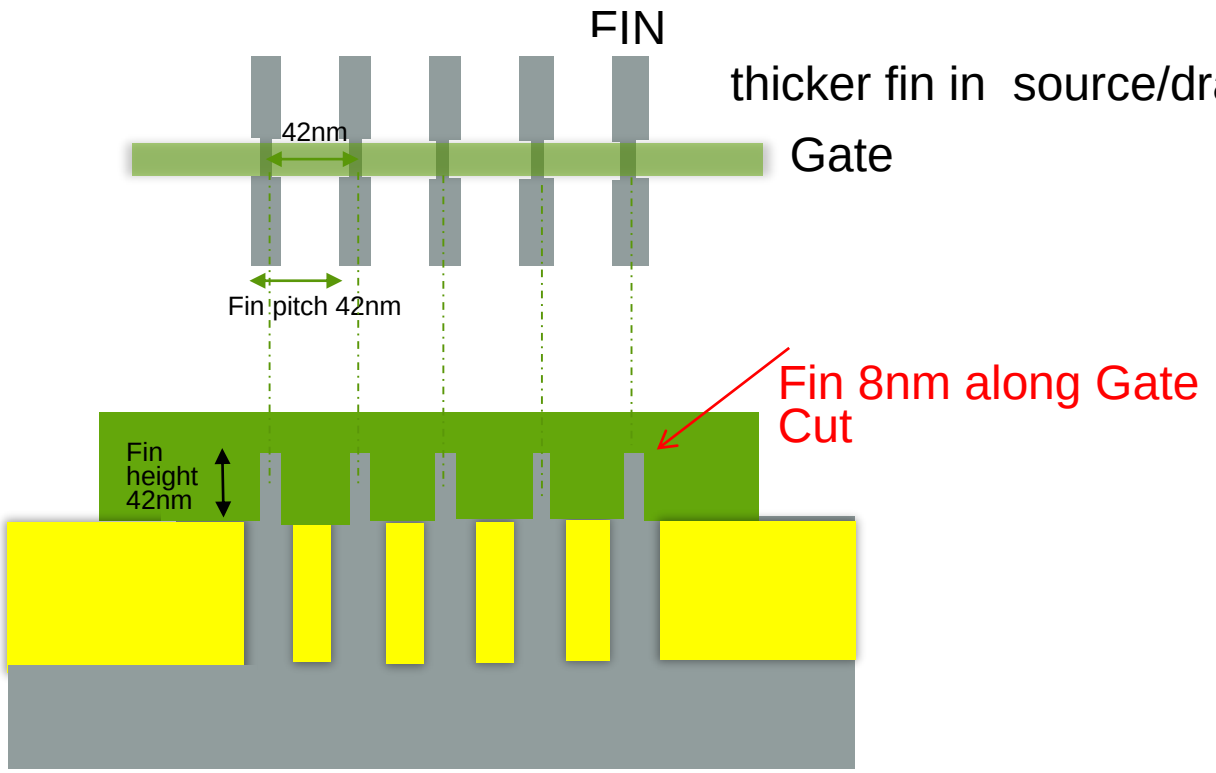
Outline

- State-of-the-art for logic devices
- FinFET Basics
- Metrics for advance logic devices
- Deeper look at some advanced device concepts
  - Transistor variation / random doping effect
  - Strain I, Ge and III- channels
  - Quantum confinement
  - Drive current and relation to mobility, velocity saturation, and density of states
  - **External resistance**
  - Sub fin doping and gate all around devices
- Conclusion: How does the roadmap evolve?

Real Limiter: Parasitic Resistance



Real Fin Shape: Why?



- **Self aligned to gate Fin thinning / straighten after gate removal?**
- **Helps with mechanical support, SiGe Stress, and spacer removal**
- **Negative is higher parasitic capacitance**

Outline

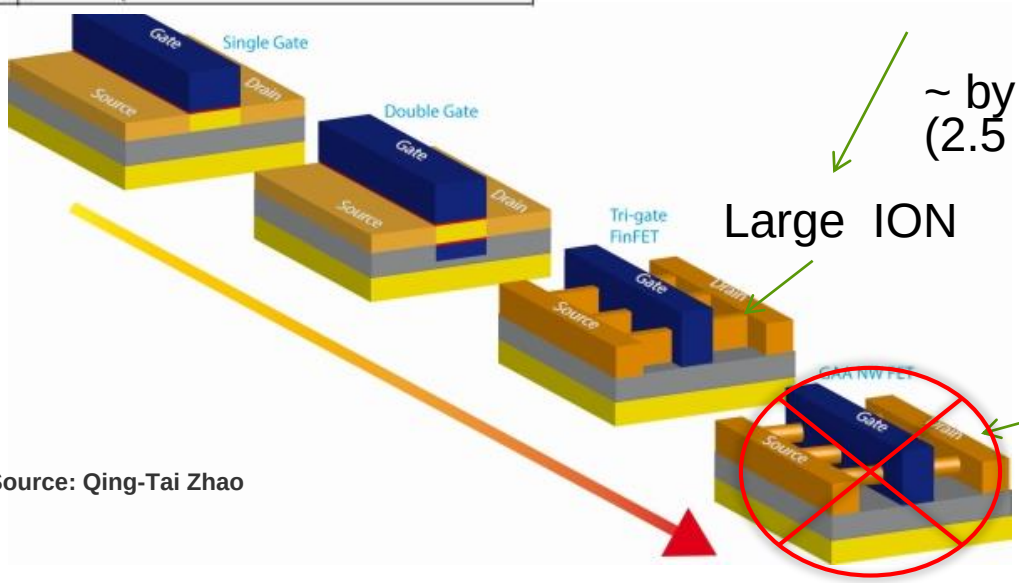
- State-of-the-art for logic devices
- FinFET Basics
- Metrics for advance logic devices
- Deeper look at some advanced device concepts
  - Drive current and relation to mobility, velocity saturation, and density of states
  - Transistor variation / random doping effect
  - Strain I, Ge and III- channels
  - Quantum confinement
  - External resistance
  - **Sub fin doping and gate all around devices**
- Conclusion: How does the roadmap evolve?

Scaling L_G for 10 and 7nm Nodes

Single gate	$\lambda = \sqrt{\frac{\varepsilon_{Si}}{\varepsilon_{ox}} t_{Si} t_{ox}}$
Double gate	$\lambda = \sqrt{\frac{\varepsilon_{Si}}{2\varepsilon_{ox}} t_{Si} t_{ox}}$
GAA structure	$\lambda = \sqrt{\frac{2\varepsilon_{Si} t_{Si}^2 ln\left(1 + \frac{2t_{ox}}{t_{Si}}\right) + \varepsilon_{ox} t_{Si}^2}{16\varepsilon_{ox}}}$

$$W_{ELECTRICAL} = FIN_{PERIMETER} = \#FINS\left(2FIN_H + FIN_W\right)$$

$$W_{ELECTRICAL} = \frac{PLANAR_{WIDTH}}{FIN_{PITCH}}\left(2FIN_H + FIN_W\right)$$



Source: Qing-Tai Zhao

~ by 10nm transistor $W_{ELECTRICAL}$ (2.5 – 3) times planar width

Large ION

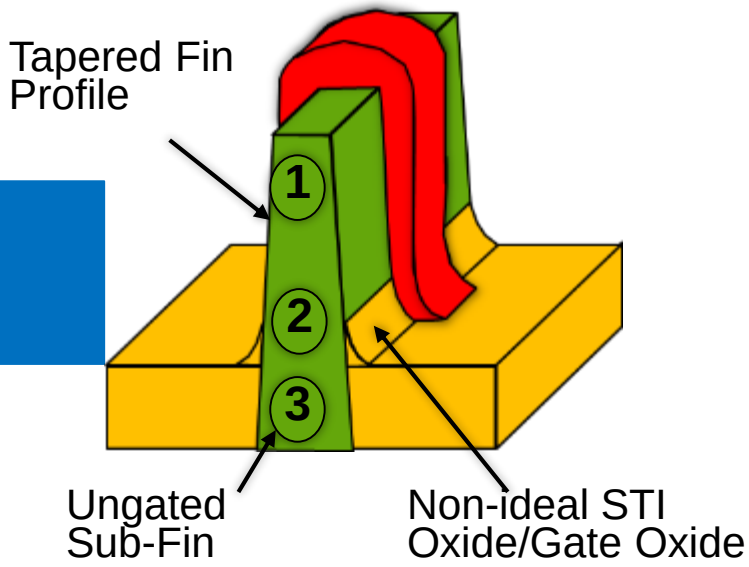
Excellent short channel control but Poor ION

- λ derived from Poisson's equation and relates to SCE and control of drain over the channel
- For strong gate control of barrier L_G > (5-10)λ

Current Double Gate Scalable to 7nm Node With 6nm Fin Thickness

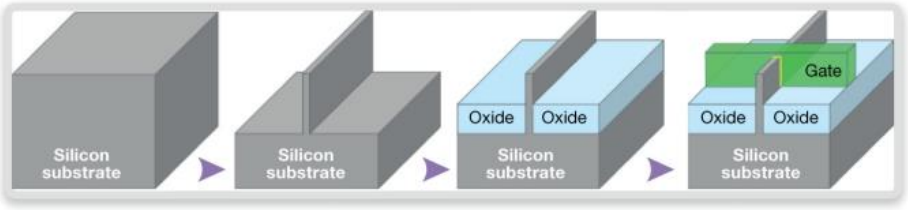
Node	22nm	14nm	10nm	7nm
Gate Pitch	90nm	70nm	~55nm	~45nm
L_G	35nm	30nm	25nm	20nm
~ L_E	25nm	20nm	15nm	15nm
t_{Si} Single Gate	9 nm	6 nm	3 nm	3 nm
t_{Si} Double Gate	18 nm	12 nm	6 nm	6 nm

- **Strong gate control of barrier L_E = 5λ**
- **Tox = ~0.9nm**
- **Will require improvements to sub-fin Leakage**

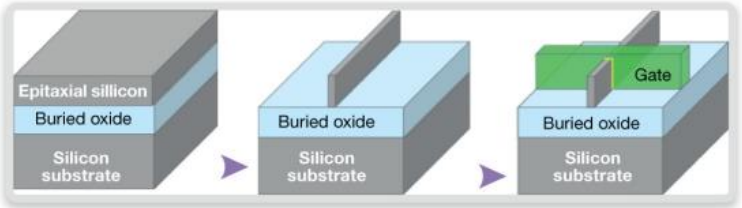


Few Ways to Address SubFin Leakage

FinFET on SOI



(a) Normal Wafer: FinFETs on regular wafers rely on a timed etch to form the fins



(b) Silicon-on-Insulator Wafer: FinFETs on SOI wafers rely on the buried oxide layer to stop the fin etch

Source: Synopsys

Bulk Sub Leakage solution

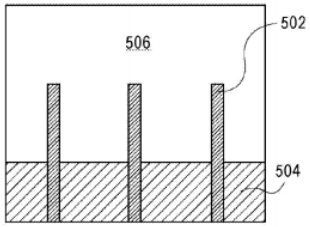


FIG. 5A

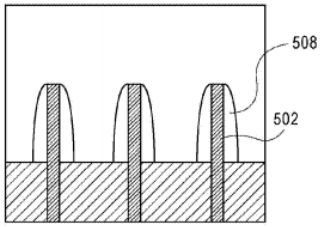


FIG. 5B

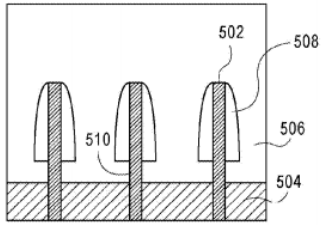


FIG. 5C

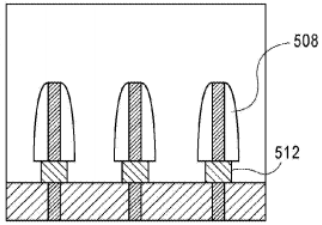


FIG. 5D

Source: Intel / US20130320455

- **Industry likely to continue with bulk FinFET for 10 and 7nm**
- **Key reason: Maintaining SiGe pFET effective mass improvement**

Gate All Arround - SiGe Stress – Larger W

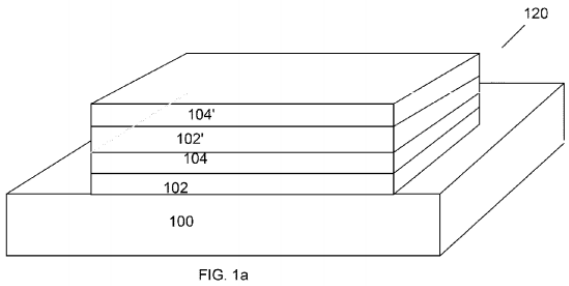
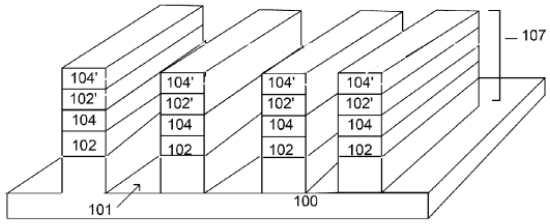
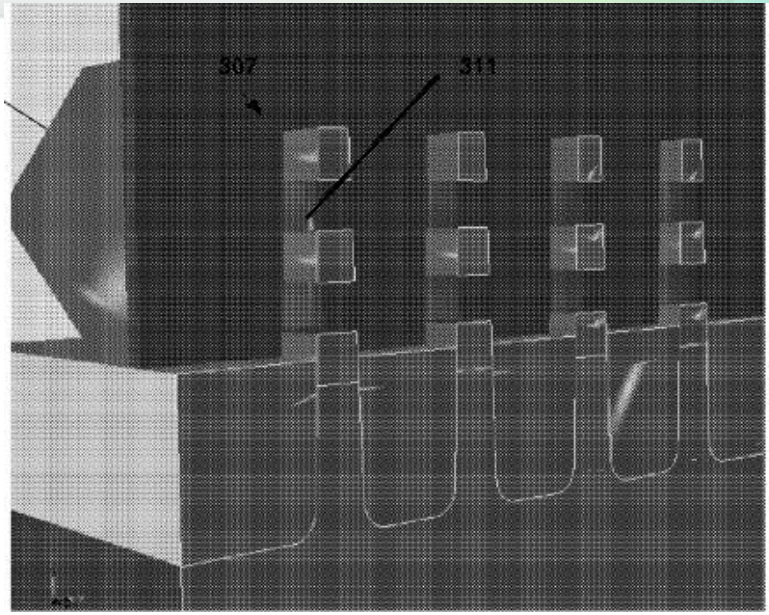


FIG. 1a

Source: Intel / US 20120138886





- No technical barrier
- Could support 5nm node
- Possible with SOI wafer technology
- Compatible with SiGe pFET effective mass improvement
- Cost is the issue (both transistor and interconnect cost with EUV required for 5nm node)

SILICON AND SILICON GERMANIUM NANOWIRE STRUCTURES

BACKGROUND

[0001] Maintaining mobility improvement and short channel control as microelectronic device dimensions scale past the 15 nm node provides a challenge in device fabrication. Nanowires used to fabricate devices provide improved short channel control. For example, silicon germanium (Si_xGe_{1-x}) nanowire channel structures (where $x<0.5$) provide mobility enhancement at respectable Eg, which is suitable for use in many conventional products which utilize higher voltage operation. Furthermore, silicon germanium (Si_xGe_{1-x}) nanowire channels (where $x>0.5$) provide mobility enhanced at lower Egs (suitable for low voltage products in the mobile/handheld domain, for example.

Conclusion

- FinFETs adopted for 14/16/22nm nodes
  - + I_{ON}, DIBL, SS, σV_T
  - - minimal L_G scaling
  - I_{ON} is from large fin perimeter
- Simple physical picture of nanoscale MOSFETs
  - Source-side injection and inversion charge sets I_{ON}
  - Mobility, scattering, velocity saturation NOT important
  - Light in-plane and heavy out-of-plane conductivity mass key
  - Strained Si still provided large improvement and enables n/p ratio ~ 1
- How future nodes 10,7, 5 nm might develop?
  - Bulk Si FinFET for 10 and 7nm
  - Tall fins
  - Subfin Leakage fix (gate all around?)
  - Semiconductors with light confinement mass and high dielectric constants (i.e. 3-5 materials) might not provide much advantage in a system with electrical and physical confinement (i.e. FinFET)